Commentary

Phonetica 2000;57:284–296

Received: November 25, 1999
Accepted: March 16, 2000

Deriving Speech from Nonspeech: A View from Ontogeny

Peter F. MacNeilage Barbara L. Davis

University of Texas at Austin, Tex., USA

Abstract

A comparison of babbling and early speech, word patterns of languages, and, in one instance, a protolanguage corpus, reveals three basic movement patterns: (1) a 'Frame' provided by the cycles of mandibular oscillation underlying the basic mouth close-open alternation of speech; this Frame appears in relatively 'pure' form in the tendency for labial consonants to co-occur with central vowels; (2) two other intracyclical consonant-vowel (CV) co-occurrence patterns sharing the alternation: coronal consonants with front vowels and dorsal consonants with back vowels; (3) an intercyclical tendency towards a labial consonant-vowel-coronal consonant (LC) sequence preference for word initiation. The first two patterns were derived from oral movement capabilities which predated speech. The Frame (1) may have evolved from ingestive cyclicities (e.g. chewing). The intracyclical consonant-vowel (CV) co-occurrence patterns involving tongue position constraints common to consonants and vowels (2) may result from the basic biomechanical property of inertia. The third pattern (LC) was a self-organizational result of pressures for interfacing cognition with action – a result which must have numerous analogs in other domains of movement organization.

Copyright © 2000 S. Karger AG, Basel

Introduction

It gives us considerable pleasure to be included in this special issue honoring Björn Lindblom. In our opinion, he has done more than any other phonetician in the 20th century to advance the cause of the discipline of phonetics. He has done this by insisting on and demonstrating the value of a fundamental conceptual framework for the discipline, summarized by his phrase 'derive language from nonlanguage' [Lindblom, 1984, p. 78]. Standing behind this advocacy is the Neo-Darwinian theory of evolution by natural selection with its fundamental tenet of descent with modification. In our discipline, Lindblom's dictum boils down to 'derive phonology from phonetics'. In his words we must derive 'the fundamental units and processes deductively from independent premises anchored in physiological and physical realities' [Lindblom, cited by

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2000 S. Karger AG, Basel

Accessible online at:
www.karger.com/journals/pho

Peter F. MacNeilage
Department of Psychology
University of Texas at Austin (B 3800)
Austin, TX 78712 (USA)

Ladefoged, 1984; see also Lindblom, 1984]. This approach is diametrically opposed to the one espoused by Ladefoged [1984] and common in phonology, according to which the most important thing about speech is that it has a level of abstract form which is largely independent of phonetics [e.g. Anderson, 1981; Halle, 1990; Kenstowicz, 1994; and a number of papers in Goldsmith, 1995].

Most phoneticians do not even take sides on these issues. Instead, they focus on the most limited of the four questions that the Nobel laureate Tinbergen [1952] tells us must be answered in order to understand a communication system: 'How does it work?' Scant attention is given to the other three questions which must be included if we are ever to explain speech in terms of Mayr's [1984] 'ultimate causes'. They are: What does it do for the organism? How did it get that way in ontogeny? How did it get that way in phylogeny?

Deriving Speech From Nonspeech

Our aim in this paper is to provide evidence in support of Lindblom's belief that the most profitable approach for phonetics is to derive speech from nonspeech. We specifically concern ourselves with Tinbergen's [1952], questions 3 and 4, ontogeny and phylogeny. Our main conclusion is that aspects of the structure of speech held in common between infants and modern languages were probably the building blocks of the first spoken words.

A little over a decade ago we began to study the acquisition of speech. Although we did not know of Tinbergen's framework at the time, we began to consider all of his questions simultaneously. The 'how does it work' question was addressed by attempting to infer the nature of movement control of the production system from acoustic information. We assumed, with Lindblom, that what speech does for the organism is that it allows us to send and receive a large linguistic message set by developing a system which responds to pressures for optimization across the conflicting constraints of ease of production and perceptual distinctiveness.

The core of a possible answer to the question 'how did it get that way in phylogeny' was provided by the Frame/Content theory of evolution of speech, initially presented by MacNeilage et al. in 1984 and 1985. According to this theory, an initial frame for speech in the form of a close-open mouth cycle produced by mandibular oscillation was subsequently elaborated by programming of different consonants and vowels for the closing and opening phases. The initial stage was considered to be dominated by motor constraints with subsequent developments highly influenced by the need for perceptual distinctiveness between message variants.

The empirical aspect of our research program involves ontogeny. The proposed answer to the question of 'how did it get that way in ontogeny' is that ontogeny recapitulates phylogeny in the sense of beginning with a similar set of motor constraints to those of early phylogeny, and then progressively acquiring the sounds and sound patterns that were introduced into languages later [MacNeilage and Davis, in press a].

There was one additional perspective with which we began this work. It was provided by Karl Lashley [1951] in a classic paper entitled 'The Problem of Serial Order in Behavior'. The problem he posed was: 'How is any sequence of actions organized?' He considered speech to be at once the most challenging and potentially the most revealing serially ordered behavior in living forms.

The Frame/Content Theory of Evolution of Speech

The Frame/Content theory [see MacNeilage, 1998, for a current version] was initially formulated as a possible explanation of the serial organization of adult speech. The key observation that led to the theory was that in segmental exchange errors, the segments almost always go into the same position in syllable structure that they came out of [see Levelt, 1989]. Initial consonants exchange with initial consonants ('tall ships' → 'shawl tips'), vowels exchange with vowels ('ad hoc' → 'odd hack') and final consonants exchange with final consonants ('top shelf' → 'toff shelp'). Phenomena such as these led Levelt [1992, p. 10] to conclude that 'probably the most fundamental insight from modern speech error research is that a word's skeleton or frame and its segmental content are independently generated'. The basic assumption of the Frame/Content theory is that the premotor frame constraining speech errors evolved from a motor frame of mandibular oscillation. The capability of inserting independently controlled segmental content elements is considered to have evolved later.

Frame Dominance in Babbling and Early Speech

One of the most salient characteristics of babbling and early speech is that the prototypical event is a relatively rhythmic alternation between a closed and open mouth configuration (e.g. [baba]). Our initial assumption was that this oscillation was related to the one we postulated as the motor frame for the earliest speech of hominids. We assumed that acquisition of speech production was a matter of 'Frames, then Content' [MacNeilage and Davis, 1990a]. Our primary interest was to determine how segmental content came to be differentiated from frames, or, in other words, how segmental independence developed [MacNeilage and Davis, 1990b].

The methodological approach we have taken is very simple, but powerful. We generate extremely large databases of phonetically transcribed babbling episodes and early words (and more recently words in dictionaries). We then evaluate the frequencies of serial organization patterns against frequencies to be expected on the basis of the overall occurrence of the individual elements in the corpus.

So far, our work on speech acquisition has been confined to the babbling stage (7–12 months), and the subsequent stage of production of single words (12–18 months). When we began the work it was already known that these two stages involved very similar output forms [e.g. Vihman et al., 1985]. We have found very little evidence that infants have developed any segmental independence by the end of the single-word stage. Instead, what we found can be summarized by the term 'Frame Dominance' [Davis and MacNeilage, 1995]. Most of the variance in vocal output from 7 to 18 months of age can be attributed to mandibular oscillation alone, with very little evidence that any of the other articulators – lips, tongue, soft palate – are moved independently during vocal episodes. We will now describe the studies that led to this conclusion.

The first problem we encountered was that there was very little knowledge about vowels in early speech acquisition. As vowels and consonants alternate with each other in babbling and early speech, both the problem of segmental independence and the more general serial order problem crucially involve vowels as well as consonants, and, specifically, the relations between the two elements. Consequently, consistent with our methodological strictures, our first study was an extremely large-scale case study of phonetic transcription of words of a single infant at the early word stage, 14–20 months, with the intention of considering vowels and vowel-consonant relations [Davis and MacNeilage, 1990].

The results of this study gave us our first inkling of the presence of Frame Dominance. We found three consonant-vowel co-occurrence patterns: coronals with front vowels, dorsals with back vowels and labials with central vowels. The two lingual patterns – coronal-front and dorsal-back – suggested that the tongue was typically not moving independently of the mandible. For the labial-central pattern, it seemed possible that lip contact was simply the result of mandibular oscillation and that the tongue was in a rest position in the center of the mouth. Consequently, apart from the positioning of the tongue in the anterior or posterior part of the mouth, which apparently occurred by the onset of the utterance, mandibular oscillation alone could be the source of all three of these patterns.

As we had no reason to believe at that time that these patterns were also present in the English language, and therefore could have been learned by the time of the first words, we hypothesized that they must be basic to early vocalization and must therefore be present from the onset of babbling. Subsequent studies of CV co-occurrences in babbling and early speech, in which we attempted to confirm these hypotheses, focussed on the co-occurrence of vowels with stop consonants, nasals and glides, because these segment types constituted the overwhelming majority of nonvocalic sounds (92% in the first-word period).

In three studies of babbling, in a total of 9 infants, all three predictions have been strongly confirmed. There were 27 possible tests of the three hypotheses (9 subjects, 3 per subject). Three could not be made because there were insufficient numbers of dorsal consonants. Of the remaining 24 tests, 23 were positive and there was 1 null result. Overall the three co-occurrence patterns were almost 30% higher than would have been expected on the basis of the frequencies of the two components of the CV pair in the overall corpus. Median ratios of observed-to expected co-occurrences in the study of 6 subjects were: coronal-front 1.28; dorsal-back 1.22, and labial-central 1.34.

A very similar result was found in two studies of early words, involving 10 infants and the same methodology [MacNeilage et al., 1997, and unpublished observations]. A total of 27 of the 28 possible tests were positive, with 1 negative outcome. The median ratios of observed-to expected co-occurrences in the 10 infants were: coronal-front 1.45; dorsal-back 1.36, labial-central 1.28. We believe that most of the null findings and counterexamples reported in other studies [Boysson-Bardies, 1993; Oller and Steffans, 1993; Tyler and Langsdale, 1996; Vihman, 1992] are a result of four methodological differences between their studies and ours. These studies uniformly involved much smaller databases than ours, sometimes divided their data into small subsets for analysis (thus reducing sample sizes), sometimes used different vowel classifications, and sometimes did not take both vowel and consonant frequencies into account in computing expected values.

Another prediction we made after the initial study [MacNeilage and Davis, 1990a] concerned a possible lack of independence between successive syllables in addition to the lack of independence we had found between successive sounds within syllables. Some background is needed in order to understand why this prediction was made. Earlier work had led to the claim that the initial intersyllabic pattern in babbling was one of syllable reduplication (repetition) and that variegation (nonrepetition) only became prominent in later babbling [e.g. Oller, 1980]. This claim seemed plausible because it was consistent with the commonsense assumption that there was a progression towards greater serial complexity as infants got older. However, subsequent results suggested that reduplication and variegation coexist in more or less equal quantities from the beginning of babbling [Smith et al., 1989; Mitchell and Kent, 1990]. We subsequently found, in a study of babbling in 6 infants [Davis and MacNeilage, 1995], that there was the same amount of variegation in the first half of the babbling period as in the second. These results cast some doubt on the conclusion that variegation involves more segmental independence than reduplication.

This finding encouraged us to hypothesize that, as in intrasyllabic organization, most of the variance in intersyllabic organization might involve the vertical, or close-open dimension rather than the horizontal, or front-back dimension of the oral cavity, and might therefore be primarily due to 'Frame Modulation' [MacNeilage and Davis, 1990a]. Consonants might vary primarily in manner of articulation, which mostly involves the amount of constriction in the tract (related to mouth closing), rather than in place of articulation, which involves changes in the front-back dimension. Vowels might vary more in height (related to mouth opening) than in the front-back axis. The mandible might be the primary contributor to both the amount of mouth closing for consonants and the amount of opening for vowels.

Both of these predictions were also strongly confirmed for 6 infants in babbling [Davis and MacNeilage, 1995] and 10 infants in first words. All infants showed both predicted effects in babbling at highly significant levels. The median ratio of observed to expected occurrences for consonants was 2.80. For vowels it was 1.42. As to speech, because some infants produced too few instances of consonant variegation in their first words for separate tests, the data for the 10 infants was pooled and the expected effect was found to be highly significant. All infants but 1 showed the vowel effect. The median ratio of observed to expected occurrences was 7.0.

The confirmation of the 5 predictions, 3 for CV co-occurrences and 2 for frame modulation, provided strong support for the Frame Dominance concept. Most of the variance in babbling, and early speech, is the result of mandibular oscillation, with other articulators tending to adopt a static configuration throughout the utterance.

Three other studies provided further support for the Frame Dominance conception. The studies described earlier only involved the frequently occurring nonvocalic sounds – stop consonants, nasals and glides. However an additional study of infrequent consonants in babbling – fricatives, affricates and liquids – showed that they are for the most part subject to the same CV co-occurrence patterns and variegation patterns as the more frequently occurring consonants [Gildersleeve-Neumann et al., in press]. An acoustic analysis showed that the vowels intervening between successive nasal consonants (e.g. [m<u>a</u>m]) were strongly nasalized, suggesting that the soft palate – like the tongue and the lips – typically stayed in the same position for entire nasalized utterances [Matyear et al., 1997]. A separate study was made of consonants that occur in absolute-final position in babbling, as earlier studies suggested that they tended to be different than consonants elsewhere in being more often voiceless and/or fricatives [Redford et al., 1997]. We found that, regardless of their voicing and manner of articulation, these consonants typically agreed with the consonant preceding them in place of articulation, suggesting that they result primarily from frame reiteration.

CV Co-Occurrence Patterns in Languages

The strength of the CV co-occurrence patterns in infants led us to ask whether this lack of segmental independence was a development-specific phenomenon or whether it was also present in adult speech. We tabulated CV co-occurrences involving stops and nasals in 12,630 words derived from dictionary counts of ten languages including several major language families – English, Estonian, French, German, Hebrew, Japanese, New Zealand Maori, Ecuadorian Quichua, Spanish and Swahili [MacNeilage et al., 2000]. We found evidence for all three patterns at an average of 18% above expectation. Seven languages showed the coronal-front pattern with a mean observed-to-expected ratio of 1.16. Eight languages showed the dorsal-back pattern with a mean observed-to-expected ratio of 1.27. Seven languages showed the labial-central pattern with a mean ratio of 1.10.

We believe that the joint occurrence of these three patterns in infants and languages indicates that they were present in the first language/s. We believe this because the biomechanical contingencies involved in these patterns are so basic. The lingual co-occurrences (coronal-front and dorsal-back) must simply be due to a biomechanical constraint on the amount of tongue movement that can readily be made between a consonant involving the tongue and a vowel that follows it. The labial-central pattern is of particular interest from the standpoint of the Frame/Content theory. According to the theory, the basic frame involves mandibular oscillation alone with no necessary active intrasyllabic movement of any other articulator. This is exactly what seems to be occurring in instances of labial-central co-occurrence in infants. The tongue is presumably in a resting position in the center of the mouth. Even though the tongue and lips are not passive in modern adult labial-central pairings [MacNeilage and Sholes, 1964] modern languages appear to have retained a tendency to preserve the position the tongue had when it was a passive accompaniment of ancestral 'pure' frame production, i.e., mandibular oscillation alone.

Early Speech: The Labial-Coronal Sequence Pattern

Much research has shown that the output forms of babbling and early speech are very similar [see MacNeilage, 1997, for a summary]. Although similarity is the mode, we have found three differences. First, labial consonants increase in first words [MacNeilage et al., 1997; see also Boysson-Bardies et al., 1992]. This seems to be a regression towards pure frames induced by the increase in functional load resulting from the task of interfacing the motor system with the mental lexicon [MacNeilage, 2000]. Second, there is more ability to vary the identity of the vowel in utterance-final position. The third trend is especially significant as it is the only clear-cut case of an increase in segmental independence that we have found during the first-word stage. The trend is called 'Fronting' [Ingram, 1974]. The first consonant in a word tends to have a more anterior place of articulation than the second. As dorsal consonants are often rare in first words, the main trend is to begin with a labial consonant, and, after the vowel, continue with a coronal (the LC pattern). This pattern was about 2.5 times more frequent than the reverse pattern in our group of 10 infants in the first-word stage [MacNeilage et al., 1999]. We also found it in nine of the ten languages we studied, occurring 2.25 times as often as the reverse pattern [MacNeilage et al., 1999].

As in the case of the increase in labials, we have interpreted this fronting as an ease-related response to the problem of interfacing output with use of the mental lexicon [MacNeilage et al., 2000]. There are two reasons for this interpretation, in addition to the implications of the labial regression effect itself. First, as discussed earlier, labials may be a simple outcome of the mandibular oscillation, at least in infants, while coronals involve an additional movement – of the tongue. Second, infants whose babbling and early speech attempts have been prevented by a tracheostomy strongly prefer labials in their posttracheostomy vocalizations [e.g. Locke and Pearson, 1990].

Why would there be an easy beginning rather than an easy ending? The existence of a separate neural system for movement initiation in vertebrates [e.g. Loeb, 1987] can be taken to mean that movement initiation poses a unique control problem, which is perhaps reduced when movement can be initiated in an easy manner.

The labial-coronal sequence effect may be a self-organizational response to the problem of simulating the serial output complexity that the infant discerns in the ambient language. The likelihood that the labial-coronal sequence is not simply the result of a copying operation in infants is suggested by the fact that it is somewhat stronger in infants from its onset than it is in adults. In addition, infants have been reported to produce this pattern even when the target word has the opposite pattern [as in 'pot' for 'top', Macken, 1978; Jaeger, 1997]. As with the CV co-occurrence patterns, the presence of this pattern in languages as well as infants suggests that it may have first developed early in language evolution. As in infants, it may have first developed in a self-organizational manner, because it was relatively readily producible, but once produced, was retained as a sound pattern for a new lexical unit.

If this interpretation is correct, the onset of the LC effect is an extremely momentous event in both evolution and acquisition, not so much in itself, but in its consequences. In both cases, it results in a quantum jump in serial complexity which, if occurring in the context of labial-labial and coronal-coronal sequences, as we have portrayed it, increases the possible disyllabic patterns by 50% at a single stroke. By making a tongue movement after the first frame, instead of prepositioning it before the utterance begins, an output discontinuity is induced. Perhaps an additional step consisted of being able to control the intervening vowel to make it either of the type that usually goes with the preceding consonant, or with the following one (central or front), thus producing a further quantum jump in sequence possibilities. Such events could rapidly conspire to give the appearance of a literally systematic discontinuity in structure of the communication system when the system is viewed from a distance, so to speak. Thus the now unique phenomenon we know as speech may have been born. However the uniqueness may lie more in the end result as we see it than in any single formative event. As suggested by Gould [1977, p. 409], 'external discontinuity may well be inherent in underlying continuity provided that a system displays enough complexity'.

Serial Organization in a Protolanguage Corpus

Our hypotheses about the form of early language/s – that they had frames, CV co-occurrence patterns and the LC sequence effect – cannot be directly tested. However we have recently found evidence for these patterns in an analysis of a 27-word corpus of putative protowords – words with common sound patterns across many existing language families – presented by Bengtson and Ruhlen [1994; MacNeilage and Davis, in press b]. In this corpus, totalling 49 syllables, there were either only 2 exceptions to the tendency to alterate between a single consonant and a single vowel, or 4, depending on whether the [ts] sequence is classified as an affricate or as a consonant cluster. The observed-to-expected ratios for the three hypothesized CV co-occurrence patterns were: coronal-front 1.94, dorsal-back 1.63, and labial-central 1.31. None of the other six observed-to-expected ratios exceeded 0.94. A chi square analysis of the overall distribution of CV co-occurrences was significant (chi square $n = 46$, $d.f. = 4$, 9.63, $p < 0.05$). The most frequent variegated consonant sequence was labial-coronal (8) and there was only 1 coronal-labial sequence.

Implications for the Frame/Content Theory

The evidence presented here suggests to us that ancestral hominids were like modern infants in going through two stages in the evolution of true speech: an initial Frame stage and a subsequent Frame/Content stage. In the Frame stage, which begins in infants at the beginning of babbling, a systematic pairing of phonation and mandibular oscillation allows sustained voiced alternation between consonants and vowels. Pure frames are the simplest result, but static nonrest positioning of the soft palate and tongue is also possible, allowing nonnasal sound production and fronted (coronal-front) and backed (dorsal-back) frames, all tending to be sustained through an entire utterance.

One other probable property of this frame stage suggests itself: babbled utterances [Davis and MacNeilage, 1995], first words, words of languages [Bell and Hooper, 1978] and the putative protolanguage forms [MacNeilage and Davis, 2000] all tend to begin with a consonant and end with a vowel. This property is probably a response to a problem of serial order in vocalization, a problem held in common with other mammals. Presumably mammalian calls in general show influences of the need to co-ordinate phonation with the departure of the articulatory system from a resting or vegetative configuration at the beginning of the call and the return to this configuration at the end of the call. In the case of hominids the problem is to sandwich one form in particular – the frame – between the prior and following vegetative states of the production apparatus [MacNeilage and Davis, in press b].

We share Lindblom's view that modern speech results from a compromise between articulatory ease and perceptual distinctiveness. The topic of this paper could have been called 'foundations of articulatory ease' except for the problem of an acceptable definition of the word 'ease'. Our conviction is that perceptual distinctiveness played/plays only a minor role in the frame stage of phylogeny and ontogeny. For the early hominid, distinctiveness was unlikely to have been a problem because initially there was only a small vocabulary, and it is easier to distinguish between a small number of possible signals (e.g. Morse code).

It is clear that the complexity of the signal level of speech must have vastly increased in evolution in order for modern languages to directly encode so many message variants. Every language has developed its modern sound and sound pattern repertoire by a historical process of paradigmatic (sound inventory) and syntagmatic (serial organization) expansion. Selection pressures have forced the expansion of the message set, even though this expansion involves the use of new sounds and patterns which take the production system out of its most comfortable range. The finding of Lindblom and Maddieson [1988] that languages with large inventories have a disproportionately large number of consonants which are difficult to produce is evidence of this process of expansion.

We suggest that these developments occurred in a second, Frame/Content stage of evolution. In our opinion, speech acquisition involves recapitulation of this sequence of events, but by progressively more precise simulation of ambient language models rather than, as in phylogeny, by a sequence of inventions of new lexicon-sound links. The LC effect may have been the first systematic syntagmatic trend in the direction of increased complexity in evolution, as it is in acquisition. In English, the long course of acquisition of fricatives and liquids is an indication of the difficulty of paradigmatic aspects of the historical process of overall repertoire expansion, and the difficulty of acquisition of consonant clusters is an indication of the difficulty of syntagmatic aspects of the process.

In the course of these developments, infants, and presumably early hominids, are/were forced to abandon some of their most basic motor patterns, in the transition to adult forms in the case of infants, and in the historical transition to modern forms in the case of hominids. These patterns, though basic to the hominid production system, are superceded because they have become incompatible with complex high speed modern speech transmission and perception. The main example of this is the virtual loss of syllable reduplication in languages, even though it is the main form of intersyllabic organization in infants, and, according to the Frame/Content theory, was also the main original form of intersyllabic organization in earlier hominids. In our study of ten languages we found that consonant reduplication occurs at only 67% of chance, levels suggesting an active prohibition of the form rather than a simple reduction to chance levels [MacNeilage et al., 2000]. Infants are therefore forced to abandon their early preference for syllable reduplication in the course of speech acquisition. It is perhaps worth noting that the disappearance of a universal and therefore (many would say) genetically specified property of speech during ontogeny is at present inexplicable within generative phonology [Drachman, 1976; Pater, 1997].

Conclusions

How successful have we been in our attempt to derive speech from nonspeech with an approach centered on speech ontogeny? How our success is judged depends primarily on the plausibility of the assumption that sounds and patterns common to infants, languages and protolanguage corpora were probably present in first words. Basic motor properties were the common element in all the patterns we found. The original form of the most basic of these patterns, the frame, may have been present since the origin of mammals, a fifth of a billion years ago [Radinsky, 1987]. Pure frames may reflect the simplest operational form of the frame. It could be said that the acoustic correlates of the resting position of the tongue in pure frames only began to signal a 'vowel' when they began to make a difference in message transmission depending on their presence or absence. In linguistic parlance this only happened when one acoustic constellation formed a minimal pair with another in the service of sending two messages. In this sense, vowel-related acoustic packages preceded vowels, and the same could be said for consonants. Speech evolved from nonspeech.

We have argued that lingual frames result from one of the most basic properties of matter of any kind, the property of inertia. We would like to turn the most obvious question around in this context. Rather than asking why the tongue does not move actively from one segment to another in infants in particular, we would ask why it should. Presumably the view that there is a genetic specification of distinctive features (usually separate features for vowels and consonants) and preferences in modes of interaction between these features would lead to the expectation of independent tongue movements from segment to segment in infants. But that view is only obtained by reification of adult sequential patterns, which are endpoints of phylogenetic and ontogenetic progressions, endpoints which, for the most part, successfully disguise their lineage.

What alternative explanations exist for the CV co-occurrence patterns? There is absolutely no evidence that there is a species-specific genetic basis for intersegmental patterns we have observed. The genetic conclusion is obviously in a different realm than the view that one of the most basic properties of languages is lingual inertia. These conclusions are on opposite sides of the mind/body dichotomy. In our opinion, the tendency toward absence of intersegmental tongue movements can be attributed to inertia, in the absence of forces from any realm that are acting to overcome it, as there is no evidence of such forces. Is this not a more economical possible explanation of the lingual co-occurrence constraints than the postulation of genetically specified units and processes of Universal Grammar that have nothing to do with the prior evolution of communication? The labial-coronal sequence pattern is in our opinion a reflection, in the serial organization of words, of an interaction between basic motor system capabilities and mental representations associated with words.

There is nothing necessarily specific to speech in the labial-coronal pattern except for the actual problem space in which the pattern is evoked. To say that some property is speech-specific simply because it occurs in speech is to finesse the problem of causality. Functional load effects are commonplace in the human sciences, although performance effects of any kind are excluded from classical generative linguistics. A similar functional load effect to the one we suggest has been shown for infant speech perception at the age when the labial-coronal effect first appears. Steger and Werker [1997] have found that infants show less discrimination of fine phonetic detail when required to pair words with objects than they show in syllable discrimination tasks.

Numerous self-organization effects on action produced in the context of a mental intentional state have been demonstrated in the ontogeny of human walking [e.g. Thelen and Smith, 1994]. We have suggested that nonspeech phenomena are not only responsible for the presence of some of the speech patterns that we see today in languages but also for some that we do not see (e.g. syllable repetition), even though there are good nonspeech reasons for them to be easily producible. Languages may have below-chance levels of intersyllabic consonant repetition because of the deleterious effects of fequent repetition of the same sound in working memory in modern high speed message transmission. It is well known in studies of working memory [e.g. Conrad and Hull, 1964] that lists for recall that include spelled letters which share sounds such as ‘dee’ and ‘bee’, and ‘ell’ and ‘eff’ lead to confusion. Equally well known is the ‘repeated phonemic effect’ whereby the occurrence of two examples of the same sound in close proximity tends to induce serial ordering errors in speech production [e.g. MacKay, 1987]. However, confusion in serial organization is not necessarily a speech-specific phenomenon. It is obviously also a factor in typing [MacNeilage, 1964], a function that did not evolve, and uses a different control system than speech.

We have provided a good deal of evidence for Lindblom's [1984] contention that it is possible to derive speech phenomena from the realm of nonspeech. We have also voiced our agreement with Lindblom that there is no alternative to this endeavor if we wish to understand the origin of speech. Time will tell how far we can get with this effort, but at the moment there is no reason for pessimism. We believe we have only seen the tip of the iceberg for the particular approach that we have adopted. A comprehensive statistical analysis of properties of an infant's early speech makes it possible to understand phonological subpatterns of that infant in the context of the overall functioning of his/her system. The absence of this context has tended to result in a good deal of indeterminacy in a large number of reports that have focussed on system fragments and resulted in the formulation of ad hoc 'rules' or 'mental strategies'.

Virtually no studies have been done on statistical properties of words of languages. We have been amazed at how much common structure in infant speech and languages we have found in our initial dictionary analyses of words. To our knowledge, our statistical study of serial organization propensities in a protolanguage corpus is the first such study ever done. The outcome of a program that studies a combination of the three approaches cannot but exceed the sum of its parts. We believe that the approach described here is a vindication of Lindblom's advocacy for deriving speech from nonspeech and of Tinbergen's advocacy of a four-pronged attack on the understanding of communication.

Acknowledgment

This work was supported in part by National Institutes of Health Grant RO1 HD2773-07.

References

Anderson, S.R.: Why phonology isn't natural. Ling. Inqu. *12:* 493–539 (1981).

Bell, A.; Hooper, J.B.: Issues and evidence in syllabic phonology; in Bell, Hooper, Syllables and segments, pp. 3–22 (North-Holland, Amsterdam 1978).

Bengtson, J.D.; Ruhlen, M.: Global etymologies; in Ruhlen, On the origin of language, pp. 277–336 (Stanford University Press, Stanford 1994).

Boysson-Bardies, B.: Ontogeny of language-specific syllable production; in Boysson-Bardies, de Shonen, Jusczyk, MacNeilage, Morton, Developmental neurocognition: speech and face processing in the first year of life, pp. 353–364 (Kluwer Academic Publishers, Dordrecht 1993).

Boysson-Bardies, B.; Vihman, M.M.; Roug-Hellichius, L.; Durand, C.; Landberg, I.; Arao, F.: Material evidence of infant selection from the target language: a cross-linguistic study; in Ferguson, Menn, Stoel-Gammon, Phonological development: models, research, implications, pp. 369–392 (York Press, Timonium 1992).

Conrad, R.; Hull, A.J.: Information, acoustic confusion and memory span. Br. J. Psychol. *55:* 429–432 (1964).

Davis, B.L.; MacNeilage, P.F.: Acquisition of vowel production: a quantitative case study. J. Speech Hear. Res. *33:* 16–27 (1990).

Davis, B.L.; MacNeilage, P.F.: The articulatory basis of babbling. J. Speech Hear. Res. *38:* 1199–1211 (1995).

Drachman, G.: Child language and language change: a conjecture and some refutations; in Fisiak, Recent developments in historical phonology, pp. 123–144 (Mouton, The Hague 1976).

Gildersleeve, C.; Davis, B.L.; MacNeilage, P.F.: Production constraints in babbling: implications for fricatives, affricates and liquids. Appl. Psycholing. (in press).

Goldsmith, J.A.: The handbook of phonological theory (Blackwell, Oxford 1995).

Gould, S.J.: Ontogeny and phylogeny (Belknap, Cambridge 1977).

Halle, M.: Phonology; in Osherson, Lasnik, Language: an invitation to cognitive science, pp. 43–68 (MIT Press, Cambridge 1990).

Ingram, D.: Fronting in child phonology. J. Child Lang. *1:* 233–241 (1974).
Jaeger, J.: How to say 'Grandma': the problem of developing phonological representations. First Lang. *17:* 1–29 (1997).
Kenstowicz, M.: Phonology in generative grammar (Blackwell, Oxford 1994).
Ladefoged, P.: The limits of biological explanation in phonetics; in Cohen, van den Broecke, Abstr. 10th Int. Congr. Phonet. Sci., pp. 31–37 (Dordrecht, Foris 1984).
Lashley, K.S.: The problem of serial order in behavior; in Jeffress, Cerebral mechanisms in behavior, pp. 112–136 (Wiley, New York 1951).
Levelt, W.J.M.: Speaking: from intention to articulation (MIT Press, Cambridge 1989).
Levelt, W.J.M.: Accessing words in speech production: processes, stages and representations. Cognition *42:* 1–22 (1992).
Lindblom, B.: Can the models of evolutionary biology be applied to phonetic problems? in Cohen, van den Broecke, Proc. 10th Int. Congr. Phonet. Sci., pp. 67–81 (Dordrecht, Foris 1984).
Lindblom, B.; Maddieson, I.: Phonetic universals and consonant systems; in Hyman, Li, Language, speech and mind, pp. 62–80 (Routledge, London 1988).
Locke, J.L.; Pearson, D.: Linguistic significance of babbling: evidence from a tracheostomized infant. J. Child Lang. *17:* 1–16 (1990).
Loeb, G.E.: Motor systems; in Adelman, Encyclopedia of neuroscience, pp. 690–692 (Burkhauser, Boston 1987).
MacKay, D.G.: The organization of perception and action (Springer, New York 1987).
Macken, M.: Permitted complexity in phonological development: one child's acquisition of Spanish consonants. Lingua *44:* 219–253 (1978).
MacNeilage, P.F.: Typing errors as clues to serial order mechanisms in language behavior. Lang. Speech *7:* 144–159 (1964).
MacNeilage, P.F.: Acquisition of speech; in Hardcastle, Laver, The handbook of phonetic sciences, pp. 301–332 (Blackwell, Oxford 1997).
MacNeilage, P.F.: The Frame/Content theory of evolution of speech production. Behav. Brain Sci. *21:* 499–548 (1998).
MacNeilage, P.F.; Davis, B.L.: Acquisition of speech: frames, then content; in Jeannerod, Attention and performance XIII, pp. 453–475 (Erlbaum, Hillsdale 1990a).
MacNeilage, P.F.; Davis, B.L.: Acquisition of speech: the achievement of segmental independence; in Hardcastle, Marchal, Speech production and speech modelling, pp. 55–68 (Kluwer, Dordrecht 1990b).
MacNeilage, P.F.; Davis, B.L.: On the origin of internal structure of word forms. Science *288:* 527–531 (2000).
MacNeilage, P.F.; Davis, B.L.: Evolution of speech: the relation between ontogeny and phylogeny; in Hurford, Knight, Studdert-Kennedy, The evolutionary emergence of language (Cambridge University Press, Cambridge, in press, a).
MacNeilage, P.F.; Davis, B.L.: Evolution of the form of spoken words. Evol. Commun. (in press, b).
MacNeilage, P.F.; Davis, B.L.; Matyear, C.L.: Babbling and first words: phonetic similarities and differences. Speech Commun. *22:* 269–277 (1997).
MacNeilage, P.F.; Davis, B.L.; Kinney, A.; Matyear, C.L.: Origin of serial output complexity in speech. Psychol. Sci. *10:* 459–460 (1999).
MacNeilage, P.F.; Davis, B.L.; Kinney, A.; Matyear, C.L.: The motor core of speech: a comparison of serial organization patterns in infants and languages. Child Dev. *71:* 153–163 (2000).
MacNeilage, P.F.; Sholes, G.N.: An electromyographic study of the tongue during vowel production. J. Speech Hear. Res. *7:* 209–232 (1964).
MacNeilage, P.F.; Studdert-Kennedy, M.G.; Lindblom, B.: Functional precursors to language and its lateralization. Am. J. Physiol. B (Reg. integ. comp. Physiol. 15) *246:* 912–915 (1984).
MacNeilage, P.F.; Studdert-Kennedy, M.G.; Lindblom, B.: Planning and production of speech: an overview; in Lauter, Planning and production of speech in normally hearing and deaf people, pp. 15–22, ASHA Rep. (American Speech and Hearing Association, Washington 1985).
Matyear, C.L.; MacNeilage, P.F.; Davis, B.L.: Nasalization of vowels in nasal environments in babbling: evidence of frame dominance. Phonetica *55:* 1–17 (1997).
Mayr, E.: The history of biological thought (Belknap, Cambridge 1984).
Mitchell, P.; Kent, R.: Phonetic variation in multisyllabic babbling. J. Child Lang. *17:* 247–265 (1990).
Oller, D.K.: The emergence of the sounds of speech in infancy; in Yeni-Komshian, Kavanagh, Ferguson, Child phonology, vol. 1: Production, pp. 29–42 (Academic Press, New York 1980).
Oller, D.K.; Steffans, M.L.: Syllables and segments in infant vocalizations and young child speech; in Yavas, First and second language phonology, pp. 45–62 (Singular Press, San Diego 1993).
Pater, J.: Minimal violation and phonological development. Lang. Acquis. *6:* 201–253 (1997).
Radinsky, L.B.: The evolution of vertebrate design (University of Chicago Press, Chicago 1987).
Redford, M.L.; MacNeilage, P.F.; Davis, B.L.: Perceptual and motor influences on final consonant inventories in babbling. Phonetica *54:* 172–186 (1997).
Smith, B.L.; Brown-Sweeney, S.; Stoel-Gammon, C.: A quantitative analysis of reduplicated and variegated babbling. First Lang. *17:* 147–153 (1989).
Steger, C.L.; Werker, J.F.: Infants listen for more phonetic detail in speech perception than in word learning tasks. Nature *388:* 381–382 (1997).

Thelen, E.; Smith, L.: A dynamic systems approach to the development of cognition and action (MIT Press, Cambridge 1994).

Tinbergen, N.: Derived activities: their causation, biological significance, origin and emancipation during evolution. Q. Rev. Biol. *27:* (1952).

Tyler, A.A.; Langsdale, T.E.: Consonant-vowel interactions in early phonological development. First Lang. *16:* 159–191 (1996).

Vihman, M.M.: Early syllables and the construction of phonology; in Ferguson, Menn, Stoel-Gammon, Phonological development: models, research, implications (York Press, Timonium 1992).

Vihman, M.; Macken, M.; Miller, R.; Simmons, H.; Miller, J.: From babbling to speech: a reassessment of the continuity issue. Language *60:* 397–445 (1985).