Cognitive Critique

LASHLEY'S SERIAL ORDER PROBLEM AND THE ACQUISITION/ EVOLUTION OF SPEECH

PETER F. MACNEILAGE

University of Texas, Austin, Texas

"...man has an instinctive tendency to speak, as we see in the babble of our young children." (Darwin, 1871 pp. 55-56)

ABSTRACT

The serial order problem is the problem of how any animal controls action sequences. Lashley maintained that in the case of speech the order of sounds was externally imposed on units that had, in themselves, no temporal valences. Speech errors reveal that the external source of control is a syllable structure (or frame) constraint on the placement of consonants and vowels (content elements) whereby these two segmental forms cannot occupy each others' positions in syllable structure. According to the author's frame/content theory of evolution of speech, the frame constraint evolved because the original form of speech was a consonant-vowel (CV) syllabic cyclicity involving a close (consonant) open (vowel) mouth alternation produced by mandibular oscillation. As the requisite mandibular elevation and depression involve antagonistic movements, there was no opportunity in the evolution of speech for control signals related to these two phases to get mixed up with each other. The central contention of this paper is that babbling, which is a rhythmic series of CV alternations powered by mandibular oscillation, is an innate

fixed action pattern which evolved as an ontogenetic affordance of the original frames for speech, and for their subsequent program-mability with content elements in both phylogeny and ontogeny.

INTRODUCTION

In one of the great papers in the history of neurobiology, Karl Lashley (1951) posed *The Problem of Serial Order in Behavior.* This is the problem of how *any* output sequence is organized. He addressed this problem because he regarded it as

> "…both the most important and also the most neglected problem in cerebral physiology" (p. 114). In his consid-eration of the problem he focused on language because, in his view, "…language presents in its most striking form the integrative functions that are characteristic of the cerebral cortex, and that reach their highest develop-ment in human thought processes" (p. 113).

This paper is about what is to be gained by considering the ve-hicle of language transmission, namely speech production, includ-ing its acquisition and its evolution, from the standpoint of its serial organization using the perspective provided by Lashley. The par-ticular topic has been addressed to some degree in an earlier paper (MacNeilage 1999). The present paper adds more recent evidence about the acquisition of speech production and the brain organiza-tion underlying speech production in support of the theory present-ed in the 1999 paper, the frame/content (F/C) theory of evolution of speech production (See MacNeilage, 1998, 2008 a, b. Also see a review of MacNeilage, 2008a, in this journal, by Jenkins, 2010). In addition, it adds an attempt to place the ontogenetic precursor to speech—babbling—in the perspective of ethology (the science of naturally occurring behavior), and in the perspective of the recent development of the new scientific discipline of Evo-Devo—evo-lutionary developmental biology. The basic thesis of the paper is that the presence of the babbling stage of prespeech organization offers us the key to understanding the evolution of the capacity for speech production, a capacity without which we might not ever have evolved spoken language, arguably our most important bio-logical possession.

INNATENESS AND THE UNDERSTANDING OF SPEECH

Probably the most prominent perspective regarding the understanding of speech in modern times has been provided by Noam Chomsky who has argued that speech is a result of our possession of an *innate universal grammar* thereby, in Fitch's words, putting innateness *at center stage* with respect to language (Fitch in press). This grammar, called by Chomsky a generative grammar, has two components: a syntactic or sentence structure component, and a phonological or sound pattern component. The latter component is of central concern in the present context.

The word *innate* refers, in the most general terms, to some inherent capacity. Unfortunately the term has a vexed history, being used typically with little detailed justification, and with various meanings, sometimes changing without acknowledgment within a single paper (Fitch, in press). It tends no longer to be used in mainstream biology. However, because it is a cornerstone of the highly influential discipline of modern generative linguistics, the notion of innateness needs to be addressed here. Chomsky has expressed himself on the importance of the innateness assertion in no uncertain terms:

> "To say that 'language is not innate' is to say that there is no difference between my granddaughter, a rock and a rabbit" (Chomsky 2000 p. 50).

Chomsky's innatist perspective has been taken up by modern cognitive science, (See Elman et al. 1996, for a critique) and by the new discipline of evolutionary psychology (Tooby and Cosmides 1992). His claim received a major impetus with the publication of Pinker's 1994 book, *The Language Instinct.* A centerpiece of Pinker's thesis was that there were grammar genes:

> "So for now there is suggestive evidence of grammar genes in the sense of genes whose effects seem most specific to the developments of circuits underlying parts of grammar" (p. 325).

The gene of most interest has been the FOXP2 gene. However, this gene exists in other mammals and even in birds, and its phenotypic scope in humans goes far beyond syntax. Despite two recent changes in the gene in hominins, which suggest that it has species-specific

consequences (see Ridley 2003 p. 215) we remain unable to link it with any specific linguistic phenomenon.

The main focus of claims of innateness has been on syntax; i.e. sentence structure. These claims will not receive much attention here. However, note that such claims have received very little support in the peer discussion of two recent target articles in the journal, Behavioral and Brain Sciences. One of these articles questions the presence of language universals, the main basis of support for claims of innateness (Evans and Levinson 2009) while the other questions the innateness of Chomsky's universal grammar itself (Christianson and Chater 2008).

More germane to the present topic are the claims for innateness of phonology. I have reviewed these claims elsewhere and found them wanting, and I only briefly summarize this review here (MacNeilage, 2008 pp. 225-242, 245-6). The main claim regarding phonological innateness is that the distinctive feature, the unit used for description of the subcomponents of consonants and vowels, is innate. Meilke (2007) made an analysis of 6,077 classes of sound participating in various phonological patterns in 628 language varieties. He found that even if he included all the sounds/patterns accounted for by three different postulated distinctive feature systems, each considered innate, one quarter of the patterns remained unaccounted for. He concluded that,

> "...phonological distinctive features no longer need to be assumed to be innate." (p. 197)

An even more radical conclusion than the one that features are not innate might be in order. Features might not even exist as functional entities that speakers and listeners have evolved to manipulate. This was the conclusion of Peter Ladefoged, perhaps the most important phonetician of the 20th century. After spending many years futilely trying to find, at the phonetic level, straightforward correlates of distinctive features (see MacNeilage 2008 pp. 233-5), he concluded that,

> "Phonological features are best regarded as artifacts that linguists have devised in order to describe linguistic systems." (2006, p. 12)

Beyond the distinctive feature, the aspect of sound patterns most often called innate is one associated with the syllable. Syllabic *sonority*, roughly translated as loudness, is considered to be an innate mental principle, supposedly revealed by the fact that the most

sonorous (loudest) sound in a syllable is the vowel, and sonority/ loudness then tends to decrease as the distance from the vowel of preceding or following consonants in the same syllable increases (e.g. Blevins 1995). But this pattern can be attributed to peripheral biomechanics rather than mental structure. A more open mouth results in a louder sound, and vice versa. The production of a syllable usually involves a progressive opening of the mouth until one reaches the center of the vowel, followed by a progressive closure from then on. So, naturally, vowels will be louder and loudness will tend to decrease as the distance of a given consonant from the vowel increases.

Another aspect of speech considered indicative of phonological innateness involves use of a concept of *markedness* (Prince and Smolensky, 1997). In this approach, sounds or sound patterns that are more frequent, are designated as more unmarked and vice versa. Then, via circular reasoning, the sounds and patterns are deemed to be explained in terms of innate markedness.

In summary, it is presently doubtful whether any *specific* aspect of language proper is innate, and, in particular, no firm candidate for innateness emerges from a review of phonology. Paradoxically, in this paper, despite the difficulties in using the term *innate,* I will maintain that there *is* one language-related phenomenon, though not regarded as part of language proper, which is innate, and that its innateness lies in its serial organization. (I continue to use the term *innate* because of the lack of a straightforward alternative, and I will try to say exactly what I mean by it when I use it.) The language-related phenomenon I refer to is babbling. Though Darwin used a different term *(instinctive)*, this is what he believed, as we have seen. But before I present this thesis it is necessary to say more about Lashley's serial ordering problem and its relevance to speech.

THE SERIAL ORDERING PROBLEM AND SPEECH

Lashley's paper was an argument against the prevailing behavioristic view that serial ordering was produced by a stimulus-response (S-R) arrangement of

> "… chains of reflexes in which the performance of each element in the series provides the excitation of the next" (Lashley, 1951 p. 114). Lashley concluded instead that "The order must … be imposed on the motor elements by some organization other than the direct associative connections between them" (p. 115).

Lashley used examples of errors at the phonological (sound pattern) level as the main evidence for his alternative conception. In one example, he pointed out that the three basic sounds in the word *right* could not be produced by an R-S-R-S-R chain because the sounds have no intrinsic temporal valence and are therefore equally capable of occurring in the sequence *tire*. tA key aspect of Lashley's conception of the serial ordering of language was his contention that,

> " … prior to the internal or overt enunciation of the sentence, an aggregate of word units is partially activated or readied" (p. 119).

In postulating this, Lashley anticipated the concept of short term memory, a concept which only became generally accepted after George Miller (1956) published a landmark paper entitled *The magical number seven, plus or minus two: Some limits on our capacity for processing information* (current citation count: 10,711). In subsequent years a good deal of attention has been paid to processes determining the serial organization of phonological material in short term memory for speech production, and also in actual memory experiments. Many aspects of the models that have been developed are applicable to both processes. One result of this work is the development of a class of models originally called *competitive queuing models* by Houghton (1990). These are models in which the relative activation of the constituent units determines their queuing priorities, and consequently the serial ordering of output. (A brief review of the history of this work is given in Bohland et al. 2010,

pp. 1505.) At the neural level Averbeck et al. (2002, 2003) provided support for this conception in a task in which monkeys copied segments of geometric patterns. These researchers located neuronal assemblies associated with individual movements in prefrontal cortex, showing that these assemblies were simultaneously activated prior to a movement sequence. Moreover, they were able to predict the serial order of the sequence from the levels of activation of the individual assemblies.

Perhaps the most important finding regarding serial ordering in studies of spontaneous segmental (consonant and vowel) serial ordering errors is that when a speech segment is misplaced it almost always ends up in the same place in syllable structure that it originated in. Most importantly, consonants almost never end up in vowel positions in output and vice versa. This basic finding led Levelt (1992) to conclude that,

> "Probably the most fundamental insight from speech error research is that a word's skeleton or frame and its segmental content are independently generated" (p.10).

My frame/content (F/C) theory of evolution of speech production (MacNeilage 1998, 2008a) takes this insight as its point of departure. It asks the question: *How did the process of programming segmental content elements into syllable frames arise?* Specifically, what is the reason for the frame constraint on speech errors whereby consonants and vowels cannot occupy each other's positions in syllable structure? The suggested answer derives from the fact that consonants are associated with a closing movement of the mouth whereas vowels are associated with an opening movement. This alternation is associated with an elevation/depression cycle of the mandible. Moreover, a simple close/open alternation forms a consonant-vowel (CV) syllable. This is the only universal syllable type in languages, and most languages probably consist primarily of sequences of this simple alternation (Maddieson 1999).

The key initial premise of F/C theory is that the frame constraint arose phylogenetically from the fact that the movements of closing and opening the mouth are antagonistic. Use of this movement cycle probably predated the capacity to program its two phases with different segments. Consequently, there was never an opportunity in the evolution of the control program for *plans* associated with the closing phase, and plans associated with the opening phase, to get mixed up with each other (MacNeilage 1998a).

The next step in formulating F/C theory was to surmise that speech probably began simply, then increased in complexity. Unfortunately, we have no direct access to speech phylogeny. But as such a simple-to-complex sequence is observable today in speech ontogeny, it seemed worthwhile to study the nature of speech ontogeny as a possible clue to phylogeny. What soon became obvious was that the first pre-linguistic speechlike behavior, namely babbling, primarily consisted of sequences of mouth closed/open alternations, produced by mandibular oscillations, with an extremely limited capacity to vary the detailed form of the alternation in single utterances. As part of a program of work on this question with Barbara Davis we designated the mandibular cyclicities of babbling as *motor frames*, arguing that the course of speech acquisition was consequently one of *frames, then content* (MacNeilage and Davis 1990). There are good reasons to believe that babbling is innate. Let us now consider the phenomenon of babbling in greater detail, with the aim of enumerating its claims to innateness.

BABBLING

What has been called canonical babbling (Oller 1980) can be defined as

> "… one or more instances of a rhythmic alternation of a closed and open mouth, produced by a mandibular elevation/depression cycle, accompanied by vocal fold vibration, and linguistically meaningless, though giving the perceptual impression of a consonant-vowel (CV) sequence" (MacNeilage in press, a).

An example of a babbled utterance is *bababa*. Babbling tends to begin rather suddenly at about 8 months of age (van der Stelt and Koopmans-van Beinum 1966) and continues until the first words are spoken about 5 months later, and beyond.

Support for the claim that babbling is to some degree innate comes from its lack of dependence on specific experience, for the first two or three months, at least. It is considered to initially be basically the same in unimpaired infants in all language environments (Locke 1983). Although its consonant-like and vowel-like sounds tend to be ones common in languages, the possibility that they result simply from imitation is contradicted by the fact that there are some

consonants that are common in languages but are virtually absent in babbling, such as [s] as in *set* and [l] as in *let*.

An additional reason to de-emphasize the experiential basis of babbling is that it tends to primarily involve rhythmic sequences in which CV forms are iterated at regular intervals, regardless of whether the target language has a similar basic structure, (which is common) or whether it tends to vary considerably from word to word in the number of consonants between vowels. In the latter situation, as in English, the fact that the number of consonants between vowels varies considerably across syllables leads to the language not sounding as if it consists of rhythmic syllabic repetitions. Nevertheless infants growing up listening to such languages are no less rhythmic in their CV syllable sequences than are infants listening to languages which adhere more consistently to the simple CV pattern.

Exactly how rhythmic is babbling from a quantitative standpoint? CVs occur at the rate of about 3 per second. In a study in our laboratory (Dolata et al. 2008) it was found that the standard deviation of intersyllabic durations in babbling infants was 24 ms. Two thirds of the intervals thus ranged from $1/40^{th}$ of a second less than the mean CV duration, to $1/40^{th}$ of a second more. This corresponds to listener's intuitions in indicating an extremely high level of rhythmicity. In addition, there is no indication that the infant is developing this rhythmicity with increased experience during the babbling period. In this respect, babbling is quite different from the pattern observed when someone is learning a motor skill, where the rhythmicity gradually increases as the skill is mastered. This high degree of rhythmicity in babbling indicates that an extremely well organized control program is in place at its onset, and this is another reason why it should be considered innate. Not all infants begin with such rhythmicity but it is nevertheless highly typical.

This rhythmicity is an extremely significant property from the standpoint of the problem of serial order because Lashley pointed out that what he called *rhythmic systems* (known today to have their neural basis in *central pattern generators*—CPGs) constitute an important class of aids to serial organization because of their capacity to integrate widely separated strands of central neural activity.

Another aspect of Lashley's approach to the serial order problem was his extreme evolutionary conservatism. He said that,

"I am coming more and more to the conviction that the rudiments of every behavioral mechanism will be found far down in the evolutionary scale and also represented even in primitive activities of the nervous system" (p. 134). He even went as far as to say that "Analysis of the nervous mechanisms underlying order in the more primitive acts may contribute ultimately to the solution of even the physiology of logic" (Lashley 1951, p. 122).

With regard to the role of rhythm generation in the conservatism of nature, Cohen (1988) has claimed, with respect to vertebrates, that *an evolutionary continuity in a biphasic locomotory cycle of flexion and extension can be traced backward over a period of half a billion years* (MacNeilage 1998, p 502). Part of F/C theory is the claim that the relatively well formed nature of the mandibular cycle in babbling comes from the fact that it is a present day manifestation of a rhythm generator that evolved in early mammals for the control of ingestive behaviors (chewing, sucking and licking) around 200 million years ago (MacNeilage 1998). As an initial orientation, one can think of locomotor and ingestive actions as cyclicities modulated by extrinsic information (e.g. in locomotion, across uneven ground, and mastication, by the size, texture, and location of the ingested substance) whereas in speech there is a cyclicity—the motor frame— which eventually becomes modulated by the intrinsic information supplied in modern adults by the segmental content component.

Two particular conclusions from neurobiology give support for this proposal, and in doing so provide evidence that the mandibular cycle of babbling indeed has some innate basis deriving from its ingestive origins. First, Lund and Kolta (2006) consider the comparative neurobiology of brainstem circuits that control mastication, and ask, *Do they have anything to say during speech?* (p. 381). Much of the work on mastication that they review has been done on cats and monkeys. They focus, as does F/C theory, on the intrinsic rhythmical pattern" underlying mastication, which they point out is produced by a CPG. They also point out that,

"In addition to controlling motor neurons supplying the jaw, tongue and facial muscles, the CPG also modulates reflex circuits."(ibid). They conclude that, "...these brainstem circuits also participate in the control of human speech" (ibid).

Second, in a review entitled *Tongue Movements in Feeding and Speech* by Hiiemae and Palmer, (2003), the authors noted that recent functional models of the tongue have implications for the mandible because they involve the operation of what they call the hyomandibular *kinetic chain* (i.e. the jaw-hyoid-tongue complex) in both feeding and in speech. They conclude that "the cyclicities associated with speech show attributes that could argue in favor of an hypothesis which proposes that the movements of speech are a subset of those used in feeding (p. 431).

Returning now to babbling, most of what we know about it comes from studies in which the consonant and vowel-like sounds are phonetically transcribed. This approach has been criticized, particularly by Oller (2000) who characterizes it as *shoehorning* non-adult sounds into adult categories. There is some truth in this assertion. An additional problem is that inter-transcriber agreement is often not particularly high. But if one keeps these problems in mind it is still possible to find out a great deal about the overall nature of babbling using the transcription approach, particularly if one combines this procedure with more direct measurement—acoustic or articulatory. Unfortunately, there is no other single approach that comes close to it in effectiveness.

Systematic study of babbling based on phonetic transcription got off to a bad start with the extremely influential claim of Jakobson, that infants babble all the sounds of the world's languages without favor (Jakobson1941/1968). The early versatility implied by this claim made it extremely implausible on motor control grounds. The opposite of Jakobson's assertion has turned out to be true, namely that the consonant and vowel-like sounds of babbling are quite limited. Figure 1 summarizes a number of facts about the sounds and syllables of babbling. Consonants produced are primarily labial and coronal. (The square brackets indicate phonetic symbols from the International Phonetic Alphabet. See Ladefoged 1993). To quote from MacNeilage and Davis (2000) in which Figure 1 was first presented:
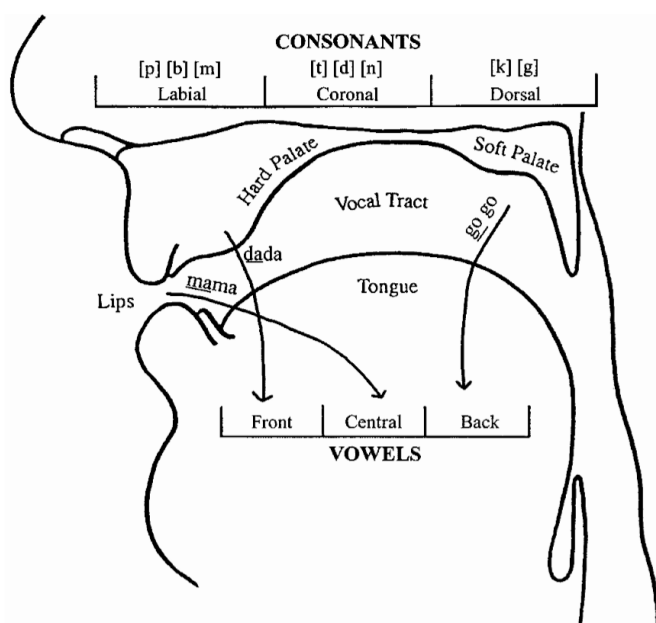
Figure 1. A schematic view of the articulatory component of the speech appara-
tus. The three arrows symbolize the three intrasyllabic consonant-vowel (CV) co-
occurrence constraints. From MacNeilage and Davis, Science, 2000, 288, p. 528.

> "The labial consonants involve lip closure and consist
> (in English) of the stop consonants that occur at the be-
> ginning of the words *pat* and *bat* and the nasal conso-
> nant at the beginning of *mat*. The coronal consonants
> involve closure in the anterior part of the mouth, (tongue
> against the hard palate) and consist of the stop conso-
> nants at the beginnings of the words *tail* and *dale*, and
> the nasal consonant at the beginning of *nail*. The dorsal
> consonants involve mouth closure in the region of the
> soft palate and consist of the stop consonants beginning
> with the words *coat* and *goat*."

Dorsal consonants are relatively rare in babbling.

In work on the relation between consonants and vowels in CV
syllables in both babbling (Davis and MacNeilage 1995) and in sub-
sequent speech, (Davis et al. 2002) we found, and other researchers
have subsequently verified, that there were three sets of preferred
relationships, indicated by the arrows in Figure 1. (In these pair-

ings, high vowels, such as the [i] in *beat* and the [u] in *boot*, were relatively rare.) Coronal consonants tended to co-occur with mid and low front vowels, such as the [ae] in the first vowel of *dada*, in the underlined example.

Dorsal consonants tended to co-occur with mid and low back vowels, such as the [o] in *gogo*, in the example in the figure. Like dorsals, back vowels are relatively rare in babbling. These two findings indicated the presence of biomechanical inertia whereby the tongue tended to stay in the same position in the front-back axis across sounds. The third pattern was a tendency for labial consonants to co-occur with mid and low central vowels, such as the [a] in *father* used in the example *baba*. This pattern also indicated an effect of biomechanical inertia, but it is perhaps a more revealing effect than in the other two findings. As the tongue is not involved in making a labial consonant, there was no mechanical constraint against it moving toward any vowel position. The fact that a central position was nevertheless preferred here too suggested that, even with no contextual constraint on its positioning, the tongue remained subject to inertia because it simply stayed in its rest position in the center of the mouth.

These results led Davis and me to suggest that babbling was subject to *frame dominance*, in that the motor frame was the main source of variance, sometimes accompanied by a preset non-resting position in the mouth—front or back—and sometimes not (Davis and MacNeilage 1995). In addition, whether or not the consonant (and the accompanying vowel) was nasal depended on whether the soft palate closed off the oronasal pathway (top right passage in Figure 1) resulting in a non-nasal delivery, or remained in its rest position.

One further point needs to be noted. There had been a common belief that while the first half of the babbling period had a reduplicative mode, whereby the same syllable was repeated, the second half of the period was characterized by variegated babbling in which successive syllables tended to differ (Oller 1980). Three studies showed that this was not the case (Smith et al. 1989; Mitchell and Kent 1990; Davis and MacNeilage 1995). In the last study (Davis and McNeilage 1995), we found, when comparing sets of two successive syllables, that the same syllable was produced about 50% of the time in both halves of the babbling period.

This finding raised another issue. The fact that 50% of pairs were variegated, even in the first half of the babbling period, suggested a cross-syllabic versatility that is somewhat at variance with our claim of frame dominance. However Davis and I have hypothesized that most of this variation might result from variation in the amplitude of the two phases of mandibular oscillation, which may not be under voluntary control, rather than from variation in tongue position. In accordance with this hypothesis, for vowels, significantly more variation was found in tongue height, which could have been produced by varition in amount of elevation of the mandible, than in the tongue front-back axis, which would require active tongue movement. For consonants, significantly more variation occurred in amount of constriction which could again have been produced by variation in elevation of the mandible, than in the place of articulation in the front-back axis, which again, would require active tongue movement (see summary in Kern and Davis 2009). These findings suggest that most intersyllabic variation might not be inconsistent with the frame dominance concept.

One interesting characteristic of babbling is that it is non-communicative. The body is often not oriented toward a listener during babbling episodes, and, in fact, babbling often occurs in the absence of a listener, as in crib soliloquies. Direct verbal attempts at inducing babbling typically result in interest, with perhaps an amused/bemused expression, but not a babbled response.

Babbling emerges from this review as a simple rhythmic universal set of CV forms with claims to an innate basis for speech. It is produced primarily by mandibular oscillation combined with vocal fold vibration. Within-utterance variability might be due primarily to adventitious variation in the amplitude of mandibular movements. Across-utterance variability in the number of CV alternations produced is simply a matter of how many iterations of the CV alternation the infant produces. The particular internal structure of babbling across utterances depends primarily on whether or not the two articulators other than the mandible—the tongue and the soft palate— are put into non-resting positions. There is only one non-resting configuration for the soft palate—elevation. For the tongue, we have somewhat arbitrarily divided the space for the non-resting tongue into two categories—coronal and dorsal— adhering to a traditional dichotomy in articulatory phonetics. However, the degrees of freedom for tongue positioning may not much exceed this two-way choice. Overall then, the simplicity of the non-mandibular

contribution to babbling consists of the use of resting positions of two articulators, and not much more than 3 non-resting positions.

IMPLICATIONS OF SIGN BABBLING

I believe there is better evidence for the innateness of the prelinguistic vocal phenomenon of babbling than there is for any aspect of language proper. However, one particular aspect of this thesis needs to be clarified. In 1991, Petitto and Marentette published a paper showing that there was an equivalent to vocal babbling in the form of sign babbling in infants with deaf signing parents. This paper quickly led to the widespread conclusion in both the field of linguistics and of cognitive science in general that humans had an innate *amodal* propensity for language. This conclusion was based on two claims, one of which is patently false, and the other of which is probably false and is presently unsubstantiated. Nevertheless, these two claims were recently promulgated by Chomsky (2006):

> "…sign languages are very much like spoken languages and follow the same developmental patterns from the babbling stage to full competence."

Instead, from the serial ordering point of view in particular, spoken language and sign language are radically different, and, regarding the timing of babbling onset in particular, there is no evidence that it is the same for vocal and for sign babbling. In fact, at present, it is not even clear what evidence regarding sign babbling would substantiate this claim. (For a more detailed treatment see MacNeilage 2008a, pp. 274-277).

It has been clear and well accepted for a long time that if one focuses on the basic superordinate units of speech and of sign, the syllable in speech, and the individual sign (which has some claim to being syllabic in rhythmic terms), speech is characterized by a succession of entities while a sign displays itself simultaneously: that is, all at once (e.g. Jakobson, 1967). The vast majority of spoken syllables involve more than one segment. Even when a syllable consists of a single vowel or consonant, segments are units of successiveness in speech. In fact, except for a few instances of *double articulations* of consonants, such as a consonant with a simultaneous labial and dorsal closure, it is impossible for two segments to occur simultaneously.

A short description of a sign, which they regard as being *identically organized* (p. 1493) to a spoken syllable, is given by Petitto and Marentette (1991):

> *"*A well-formed syllable [sign] has a handshape, a location, and a path movement (change of location) or secondary movement (change in handshape, or orientation)" (p. 1495).

These three properties are each spread across the whole sign without discrete temporal subcomponents. However, in the opinion of Brentari (2002), a leading authority on the *phonology* of sign language, one similarity between speech and sign language is that they both have Cs and Vs although,

> "Cs and Vs are realized [in sign language] at the same time rather than sequentially" (p. 60).

Brentari's criteria for concluding that signs have Cs and Vs are extremely dubious. But more importantly, from a serial ordering point of view, two entities that Brentari regards as similar could not be more different, and the difference is not just a matter of semantics. The superordinate unit of speech almost has internal segmental subcomponents with a serial order whereas the superordinate unit of sign, the sign itself, virtually never does. Lashley has given us a way of finding out about the functional organization of output in the form of error analysis, and such an analysis shows that phonological errors in speech and sign are very different (see MacNeilage in press, b, for detailed analysis).

Five types of segmental serial ordering errors of speech were identified by Shattuck-Hufnagel: exchanges (spoonerisms), substitutions, shifts, additions and omissions (Shattuck-Hufnagel 1979). By contrast, in an analysis of sign errors, Hohenberger et al. (2002) found that only exchanges and substitutions of handshapes, locations and movements occurred in sign production. Why the difference? In a more detailed look at speech errors, one finds that vowels are like the three sign parameters in only being subject to exchanges and substitution errors. These two error types could be called replacement errors, while the other three errors could be called number-changing errors. Additions increase number, and omissions decrease number, in absolute terms. Shifts change number locally, by reducing the number of segments in the immediate region that the sound is shifted from, and increasing number in the region that the

sound is shifted to. In speech only consonants are subject to such errors, as well as being subject to both types of replacement errors.

The analysis suggests that the vowel in speech is an obligatory component of the syllable and therefore cannot be changed in number. Correspondingly, the three parameters of sign are obligatory components of the manual syllable. Without them all there would not be a sign. Number changes can not occur in sign errors because there are no entities in signs comparable to the spoken consonants. The latter are in some sense optional marginal components of the spoken syllable, marginal to the obligatory vowel nucleus. This analysis adds important information to that coming from structural description in support of the conclusion that syllables and signs, rather than being *organized identically*, as Petitto and Marentette (1991) have claimed, are radically different. Thus, from a serial ordering point of view, the notion of a single innate amodal capacity underlying speech and sign production is inappropriate.

Beyond the question of structure, the other factor that has led many people to believe that speech and sign have a common amodal basis is the uncritical acceptance of the claim that speech and sign have the same chronology in terms of developmental landmarks. The three landmarks involved are babbling onset, word onset and syntax onset. Present concern is with the first landmark—babbling onset. The belief that vocal and sign babbling have the same chronology appears to have arisen from a single sentence from the paper by Petitto and Marentette (1991), on the sign babbling in two deaf infants. They assert that,

> "...by age 10 months, they were well into the syllabic
> manual babbling stage which occurred at the same time
> as in hearing infants (age 7 to 10 months)" (p.1494).

However, no data were presented in their paper to support the claim about when babbling actually began in these infants, and even if there were data, an N of two infants is insufficient to conclude that, on average, sign babbling in a population of infants exposed to sign language began at *any* typical point in time.

There is an additional problem with determining when sign babbling begins. How are signs to be differentiated from other manual phenomena? Referring to the three sign properties of handshape, location and movement, Meier and Willerman (1995) pointed out that,

> "except for statically held postures, every gesture, in-
> cluding nonlinguistic ones, will meet these criteria" (p.
> 396).

Accordingly, a large proportion of the 2,530 spontaneous manual movements of newborn infants described by Ronnqvist and von Hofsten (1994) qualify as signs, leading to the problematic conclusion that spoken and sign babbling do not have an amodal basis because sign babbling begins at birth rather than in the 3rd quarter of the first year.

Finally, perhaps to the surprise of the reader, the insistence on innateness of language by Chomsky and his supporters, either in aspects of speech or in an amodality of speech and sign, is, strictly speaking, irrelevant in the context of the present concern with the serial ordering problem. As Houghton and Hartley (1995) pointed out,

> "Theoretical linguistics … only concerns itself with the
> internal representation of serial order (*competence*) and
> not with its execution (*performance*). … In such a con-
> text, serial order will not appear to be any problem at
> all" (p. 2).

Even though the nativistic stance is therefore irrelevant, I have spent some time on it here because the scientific community tends not be aware of this fact, and consequently takes declarations regarding competence to be applicable to performance. But as we have seen, this nativism is, in fact, even more irrelevant in the present context than the competence/performance distinction makes it, because nativistic phonological concepts such as distinctive feature, syllabic sonority, markedness, and an amodal phonology lack even a potential for understanding serial ordering of speech in particular.

ACQUISITION OF THE FIRST SPOKEN WORDS OF INFANTS

Until now we have been considering babbling. Speech occurs when an infant tries to produce words. The almost exclusive source of interest in first words is in the concepts that they communicate. The most important thing to say about words in the present context is that they are produced almost exclusively by drawing on the existing babbling repertoire. It is fair to say that, at the transmission

level, words *are* babbling episodes pre-empted for communicating a concept. This fact is crucial in understanding the importance of babbling. Babbling is not just a set of throwaway vocalizations that occur prior to getting down to the real task of saying words. Babbling is, in the most literal sense, the basis of the spoken component of the first words.

One specific example of this is the fact, already mentioned, that the three CV co-occurrence preferences are just as strong in first words as in babbling. Furthermore, unpublished observations in our laboratory indicate that these preferences exist not only in correctly produced words but also in incorrectly produced words, including instances in which they are, in fact, the source of the error. For example, the mid front vowel favored with coronal consonants would be correctly produced if the infant correctly said *dead* but incorrectly produced if the infant said *dead* instead of producing the higher vowel in *did*.

More generally, most of the huge number of early word errors an infant produces can be regarded as replacements of sounds or sound patterns which are not yet in the babbling repertoire with ones that *are*. Here are some examples from MacNeilage (1997). One very common error is called final consonant deletion (e.g. [baet] (*bat)* → [bae]. Final consonants are quite rare in babbling and therefore omitted in these kinds of errors. Another common error is called consonant harmony. The same consonant occurs twice in a word when only one instance of it is called for (e.g. *doggy* → *goggy*). In babbling, a given consonant is more often followed by an identical one than by any other one, and this pattern is retained in this type of error. In the frequently encountered case of consonant substitutions, consonants not prominent in the babbling repertoire are replaced by consonants that are in the repertoire which are similar to them (e.g. *look → yook*; *rabbit → wabbit*; *sat → dat*). Another relatively common error is called unstressed syllable deletion. In this error type, the infant leaves out an unstressed syllable in a word (e.g. *banana → nana*). Babbling consists of one or more relatively long syllables, longer even than adult stressed syllables. But unstressed syllables are required to be relatively short. Consequently a relatively short syllable can therefore be considered to not be in the infant's babbling repertoire.

How then does the infant proceed to correct these errors? Perhaps the main question from the serial ordering standpoint is how does an infant develop from having a tendency to reduplicate

syllables (the canonical form of frame dominance), to a very strong tendency to variegate successive syllables common to all languages? Here is one clue as to how this begins to happen (MacNeilage in press, a):

> "There is a strong trend across language environments for the first systematic step toward intersyllabic variegation in word production to involve a preference for *starting a word with a lip consonant and following it after the vowel with a tongue-front consonant*—for example, in the word [bado] for *bottle* rather than [dabo] for *double*. … Davis and I have suggested that this particular pattern may have developed by self-organization (MacNeilage and Davis 2000). It may be a case of beginning simply with what we call a *Pure Frame*, (the lip consonant-central vowel sequence, produced only by the mandible) because, in lay terms, it is easier. But once having begun, the infant can then take the additional step of making the tongue movement needed to get the tongue-front consonant. This conjecture is consistent with evidence from neurophysiology that *starting* to move is a special problem for the brain, because it is addressed with dedicated circuitry" (Gazzaniga et al. 1999).

Obviously, a great deal more than what is covered here is involved in getting to the point where a child uses the phonological component of the lexicon of his/her native language correctly. Unfortunately, as infants acquire more sounds and more different serially ordered patterns, it becomes increasingly more difficult to make the kinds of statistical generalizations presented here about the acquisition process, even for single infants, let alone for groups of infants.

ONTOGENY AND THE NATURE OF SPEECH

What are the implications of babbling and of early speech acquisition, for the nature of speech in general, including its evolution? One possibility is that the intrasyllabic CV co-occurrence preferences, and the intersyllabic LC effect, are just growing pains, reflecting the constraints of an immature system, and are superceded as the infant matures. Quite the reverse. Our research group has shown that both the three CV preferences (MacNeilage et al. 2000) and the LC ef-

fect (MacNeilage et al. 1999) tend to be present in languages, and both of these results have been replicated by Rousset (2003). The typical presence of both of these effects in languages lead us to believe that ontogeny of speech recapitulates its phylogeny such that speech itself began with a frame stage, then subsequently evolved to the frame/content stage. Of these two stages, the frame stage, as will be discussed later, may have evolved under selection pressures associated with vocal communication even before the origin of words. But the progression into the frame/content stage in evolution must have been a result of selection of cultural transmission units—what Richard Dawkins (1976) calls *memes*—as a response to pressures to increase the size of the communicative message set by making more different word forms after the earliest words originated.

PHYLOGENY OF SPEECH

Let us now consider the phylogeny of speech in more detail. How might the frame stage of speech have evolved by descent with modification? The cyclical property of the mandible is probably as old as vertebrates. But it received an enormous boost in vertebrate evolution more than 200 million years ago with the advent of internal temperature regulation in the transition from reptiles to mammals. Maintenance of a constant body temperature required higher rates of food ingestion, and mandibular cyclicities for chewing, sucking, and licking, were adaptive responses to this requirement. I have suggested that this cyclical capability of the mandible might have been adapted to eventually form the motor frame for syllable production (MacNeilage 1998). The possibility that the progression from ingestive cyclicities to syllables had an intermediate phase is suggested by the widespread existence of various visuofacial communicative cyclicities—lipsmacks, tongue smacks and teeth chatters in other primates, all of which involve mandibular oscillation, and some of which (called *girneys* or *grunts*) are accompanied by phonation.

As to the origin of lipsmacks, Van Hooff (1967) has suggested that they may have evolved their communicative status from cyclical ingestive movements elicited during a manual-grooming event. Animals looking forward to finding a food item, such as a salt grain, in an individual instance of grooming, might have begun chewing movements in anticipation of such a discovery.

Initial selection pressures for the protosyllabic cyclical forms (in effect, lipsmacks with voicing), may have come from *vocal* grooming. According to Dunbar (1996), vocalization may have been substituted for manual tactile contact as ancestral hominin group sizes increased enough to make the latter behavior ineffective as a device for social cohesion. It is also possible that an evolving capacity to learn vocalization occurred when vocal grooming became important as part of a general-purpose mimetic capability, selected to enhance group solidarity, as suggested by Donald (1999). This capability to recreate the observed actions of others is almost as salient and unique in humans as is speech. It is evident today in movements related to music, dance, opera, ballet, movies, games, sports, etc. —behaviors not present in other living primates.

THE ETHOLOGY OF VOCAL BABBLING

An evolutionary perspective on babbling can be enhanced from a consideration of the discipline of Ethology—the science of naturally occurring behavior, historically non-human behavior. A central phenomenon in this field is the *fixed action pattern* a wide stereotyped movement or movement complex, often called *innate* as it can exhibit basic properties independent of experience. More recently these patterns have been incorporated into the broader category of *motor mechanisms* (Hogan 200l, p. 230) with less, but still some emphasis on innateness. Eibl-Eibesfelt (1989), in a discussion of these patterns, draws attention to their *form constancy* (pp.25-32). Prominent examples of such patterns, which are often extraordinarily complex, are rodent grooming, mud bathing, food catching, and courtship rituals (Fentress and Gadbois 2001). Mouse grooming, for instance involves 4 phases, all repeated several times, with some constancy in the order of phases (Berridge et al., 2005)

There are innumerable instances of such innate adaptive motor patterns, both in vertebrates and in invertebrates. Davis and Richards (2000) point out that, with respect to intentional communicative movements in particular,

> "A common and predictable feature of such intentional
> display movements is rhythmic (oscillatory) repetition."
> (p. 1)

Examples that they cite include male mallard ducks bobbing their heads up and down to a female, spiders waving their palps up and down as a form of courtship, Sceloporus lizards identifying one another from push-ups, and chimpanzees swaying from side to side as either a threat, or for courtship. One property of many of these patterns is that they manifest themselves prior to being used for an adaptive purpose. For example, the action pattern of pecking in chickens can be observed before it is used for ingesting food objects. There seems to be no reason not to include babbling as a fixed action pattern, and to conclude that it evolved as a basis for speaking. As discussed earlier, it certainly has the stereotypy, the form constancy, and the independence of specific experience that is characteristic of fixed action patterns, as well as the oscillatory character that these action patterns often possess. It also appears prior to its eventual use for communication.

It is pertinent to add the observation by Thelen (1981) that babbling is far from unique as an infant rhythmic behavior. Instead, it is simply one of a wide variety of repetitive rhythmic movements characteristic of infants in the first few months of life. These rhythmic movements include *kicking, rocking, waving, bouncing, banging, rubbing, scratching, swaying* [...] (Thelen, 1981 p. 238). She believed that such *rhythmic stereotypies are transition behavior between uncoordinated behavior and complex coordinated motor control.* In her opinion, they are *phylogenetically available to the immature infant* (p. 253), which corresponds to *innate* in the present context. In this view,

> "rhythmical patterning originating as *motor programs essential for movement control* … are 'called forth', so to speak, during the long period before full voluntary control develops to serve adaptive needs later met by goal-corrected behavior" (ibid. italics mine).

It is a very short step to conclude that the human infant cyclicities mentioned by Thelen, that involve limb movement, are offshoots of the locomotor CPG posited by Cohen (1988).

THE EVO-DEVO PERSPECTIVE AND BABBLING

When we consider the origin of babbling, the new discipline of evolutionary developmental biology (Evo-Devo) might have something to offer. Proponents of this discipline have successfully shown that changes in the roles played by regulatory genes in affecting the timing, or the strength, of expression of other genes during development, can have profound phylogenetic effects by producing differences in animal groups descending from a common ancestor. This conceptual framework goes beyond the classical genetic conception that evolution of form is directly specified by genes. Although work in this relatively new domain is mainly focused on the evolution of animal morphology—body shape and pattern— such as the structure of forelimbs during the transition from aquatic to land animals (Shubin et al. 2009), there is no reason why similar processes would not have consequences for animal *behavior*. Developmental changes, resulting in the developmental picture we see today may have led to the fixation of babbling in the human infant developmental ethogram (behavioral repertoire), with desirable consequences for adult behavior in subsequent generations.

It is instructive to consider how long ago this could have happened if, in fact, it evolved in this manner. Babbling today certainly does not appear to be a work in progress. Even though we have no good idea as to exactly when the first words were spoken, nothing about babbling suggests that it might not have been available from the very beginning of words.

As Goodman and Coughlin (2000) point out in their introduction to a special issue on developmental evolutionary biology in the Proceedings of the National Academy of Sciences:

> "Certainly the old maxim *ontogeny recapitulates phylogeny* could be the evo-devo battle cry" (p. 4425). If so, the F/C theory fits nicely within the discipline. However, they go on to suggest that "a more apt saw would be *altering ontogeny formulates new phylogeny*" (p. 4425).

If we apply this to language, an alteration in the developmental program for vocal communication in earlier hominins might have made available to them the sound patterns to be used in the first words.

AN OSCILLATORY MODEL OF MODERN ADULT SPEECH PRODUCTION

Finally, we come full circle, so to speak, by noting that Vousden et al. (2000) have presented an oscillatory-based model of the dynamics of speech production. This work follows an earlier paper by Brown and Vousden (1998) in which they characterize oscillators as *evolutionary primitive* adaptations (p. 188) and note their use in characterizing a number of human behaviors in addition to motor control—feature binding in visual processing, models of perception, variable binding in reasoning, and time estimation behavior. In that paper, they suggest, in particular, that,

> "...oscillators may underpin adaptively rational sequential foraging behavior in animals ..." (p. 188).

In their 2000 paper on speech, Vousden, Brown and Harley evaluate their model against a new database of 6,753 phoneme movement errors—exchanges, anticipations, and perseverations. (Anticipations and perseverations correspond to substitutions in Shattuck—Hufnagel's classification presented earlier. These errors are considered to be *triggered* by an instance of the same sound in the vicinity.) They find their model to be more successful than others in its ability to account not only for the syllable position effect, but also for the relative frequencies of different types of errors, movement error distance gradients, and the well known effect of phonological similarity on errors.

There are two bodies of knowledge regarding the neural organization of speech production that are particularly germane to the model of Vousden, Brown and Harley. First, Giraud et al. (2007) have identified endogenous cortical rhythms which they associate with speech-related functions. Among other things they found a 3-6 Hz power band in the lower part of the motor cortex which, in their opinion,

> "...offers a direct neural underpinning for the F/C theory of speech that assumes that syllables are phylogenetically and ontogenetically determined by natural mandibular cycles occurring at about 4 Hz." (p. 1132). They consider that overall, their findings, "...emphasize the role of common cortical oscillatory frequency bands for speech production and perception and thus provide a brain-based account for the phylogenetic emergence and shaping of speech from available neural substrate" (p 1133).

Second, I have argued that with the evolution of speech from prespeech communicative capabilities, superordinate control of the mandibular cycle, or motor frame, moved from posterior *lateral* frontal cortex to posterior *medial* frontal cortex. It came to occupy a region formerly called the Supplementary Motor Area (SMA), but now divided into Pre-SMA and SMA subcomponents (MacNeilage 1998; MacNeilage and Davis 2001). The most highly developed neurobiological model of the control of speech production, a model which includes tests of quantitative predictions has been presented by Bohland et al. (2010). They credit F/C theory for their adoption of the control dichotomy of a Frame component in medial cortex and a Content component in the lateral cortex of Broca's area and its surround.

This medial region was presumably responsible for the rhythmic CV repetitions of the same CV in Broca's patient *Tan*, and similar CV automatisms in many other global aphasics, patients who lack the use of the lateral surface of the left hemisphere, while retaining the use of posterior medial frontal cortex (MacNeilage and Davis 2001). Electrical stimulation of this region, and the presence of irritative lesions affecting this region, also result in the production of automatisms in which a single CV form is rhythmically reiterated (MacNeilage and Davis 2001). These phenomena reveal that the basic rhythmic CV alternation capability, originating ontogenetically in babbling, remains present in the brain throughout the lifespan. This medial premotor region would thus appear to be the most likely source of the oscillatory syllable-level component of the model of Vousden et al. (2000).

In conclusion, the syllable-level oscillatory mode of adult speech production that Vousden et al. (2000) suggest is the logical consequence of the present contention that the mandibular cyclicity of babbling is the main phylogenetic/ontogenetic basis of speech, and continues to be its basis throughout the life span. Complementary work at the neural level suggests that we may be converging on an integrated neurobehavioral conception of the endpoint of the evolution of speech production that Lashley might have found to be satisfying in a number of respects.

POSTSCRIPT

As suggested earlier, Darwin was probably right about babbling as he was about so many other things. As Fitch (in press) reminds us:

> "Regarding language, Darwin accepted that language is an *art* rather than a *true instinct* (because words and rules must be learned), but he also observed that language 'differs widely from all ordinary arts, for *man has an instinctive tendency to speak, as we see in the babble of our young children*; whilst no child has an instinctive tendency to brew, bake, or write.' (Darwin 1871, pp 55-56)" (Italics mine.)

I am suggesting that babbling is indeed an instinctive tendency if we replace the word *instinctive* with *innate*. But also, as Darwin seemed to believe, it may be the *only instinctive tendency* associated with language.

SUMMARY

Lashley's problem of serial order is the problem of how any animal organizes its output in the time domain. The purpose of his landmark paper was to suggest an alternative to the then prevailing chain reflex theory of serial ordering of behaviorism according to which each output unit provides the unique stimulus for the next. His focus was particularly on speech/language as he regarded it as the highest form of serially ordered behavior in the animal kingdom. His main source of evidence for his alternative thesis was segmental (consonant and vowel) serial ordering errors. He proposed that output units have no temporal valence in themselves, but are first activated a few at a time in temporary storage and then have a serial order imposed on them from another source. He also emphasized the importance of *rhythmic systems* in serial ordering and expressed an extreme evolutionary conservatism regarding the origins of the control of serial ordering.

Subsequent evidence from segmental speech errors has shown that the other source determining the ordering of erroneous outputs, and consequently also determining normal speech production, is syllable structure. Consonants and vowels do not occupy each other's positions in syllable-level output when they are erroneously misplaced in a sequence. Levelt (1992) has suggested a frame/content metaphor to characterize this process: segmental content elements are inserted into syllable structure frames.

This author has suggested a phylogenetic and ontogenetic underpinning to the frame/content mode of speech organization (MacNeilage 1998; 2008a, b). According to this frame/content (F/C) theory, the frame constraint on modern adult speech evolved because the vocal communication system that was used for the first words was a cyclical alternation between a mouth-closed (consonantal) and a physiologically antagonistic mouth-open (vocalic) configuration. As a result, there was no opportunity for the control programs associated with these antagonistic actions to become mixed up with each other in subsequent evolution.

Evidence for evolutionary priority of this *motor frame* comes from babbling, which primarily consists of a rhythmic consonant-vowel (CV) alternation produced by mandibular oscillation (e.g. *bababa*), with minimal accompanying actions of other articulators. In an ethological context, babbling can be seen as one of an enormous number of innate, ontogenetically installed fixed action patterns, used, from invertebrates onward, as the basis for adaptive actions. From the perspective of evolutionary developmental biology (Evo-Devo), babbling may have evolved as an ontogenetic adaptation facilitating hominin speech evolution.

The mandibular cyclicity underlying speech probably evolved by exaptation from the mandibular cyclicity underlying ingestive processes of chewing, sucking and licking in early mammals, as long as 200 million years ago. There was probably an intermediate stage of the kind revealed in the communicative cyclicities of lips-macks, tonguesmacks and teeth chatters all of which continue to be present in many other modern primates.

In acquisition of speech, the frame, while eventually becoming programmable with content elements, remains the basis for the serial organization process across the lifespan, as evidenced by a number of pathological circumstances, including one form of global aphasia, whereby a rhythmic sequence consisting of iterations of a single CV syllable is the only vocal output that the patient can produce. Additional neurological evidence for the centrality of the syllabic frame across the lifespan takes the form of a 4 per second (roughly syllable-rate) neural cyclicity, associated with speech production/perception, particularly notable in the brain region which includes Broca's area. Finally, Vousden et al. (2000) primarily motivated by the existence of the syllable structure constraint on adult speech output, have constructed a model of real-time speech pro-

duction as an *oscillatory* process, and their model accounts for a number of salient properties of segmental speech errors.

Chomsky has provided an alternative conception of the nature of speech as a consequence of an innate phonological component of Universal Grammar. A review of several putative innate phonological components of speech finds evidence for them to be wanting. The related hypothesis that the co-existence of spoken and signed babbling indicate an innate amodal phonological component of language is found to be without foundation.

In conclusion, the F/C theory provides a possible Neodarwinian solution to the problem of serial order posed by Lashley for the case of speech. It conforms to his strictures regarding the essential nature of the serial ordering process, vindicates his proposal regarding the importance of rhythmic systems in the control of serial ordering in the case of language, the behavior he found of most interest, and resonates with his conviction that solutions to the problem of serial ordering probably require understanding of the deep phylogenetic precursors to human functions.

The contention of the theory that was of central concern here is that the mandibular cycle of babbling (the motor frame) has evolved as an innate ontogenetic/phylogenetic basis for the serial ordering of speech, and is a time-domain precursor to the ability to program the internal structure of the frame with phonological content elements—consonants and vowels. Babbling has the stereotypy, the form constancy, and the independence of specific experience that characterizes innumerable innate fixed action patterns in nature, and also the rhythmic oscillatory character that they often possess. In addition, like many of these patterns, it appears prior to its use, in this case for communication.

REFERENCES

Averbeck BB, Chafee MV, Crowe DA, Georgopoulos AP (2002) Parallel processing of serial movements in prefrontal cortex. Proc Natl Acad Sci 99:13172-13177

Averbeck BB, Chafee MV, Crowe DA, Georgopoulos AP (2003) Neural activity in prefrontal cortex during copying of geometrical shapes. Single cells encode shape, sequence and metric preference. Exp Brain Res 150: 127-141

Berridge KC, Aldridge JW, Houchard KR, Xiaoxi Z (2005) Sequential super-stereotypy of an instinctive fixed action pattern in hyperdominergic mice: a model of obsessive-compulsive disorder and Tourette's. BMC Biology 3:4

Blevins J (1995) The syllable in phonological theory. In: Goldsmith J (ed) Handbook of phonological theory. Blackwell, Oxford, pp 206-244

Bohland JW, Bullock D, Guenther FH (2010) Neural representations and mechanisms for the performance of simple speech sequences. J Cog Neurosci 22: 1504-1529

Brentari D (2002) Modality differences in sign language phonology and morphophonemics. In: Meier RP, Cormier K, Quinto-Pozos D (eds) Modality and structure in signed and spoken languages. Cambridge University Press, Cambridge, pp 35-64

Brown GDA, Vousden JI (1998) Adaptive analysis of sequential behavior: oscillators as rational mechanisms. In: Chater N, Oaksford M (eds) Rational models of cognition. Oxford University Press, Oxford, pp 165-193

Chomsky N (2000) The architecture of language. Oxford University Press, Oxford

Chomsky N (2006) Some simple evo-devo theses: how true might they be for language. Paper presented at the Morris Symposium on the Evolution of Language, Stony Brook, NY

Christianson MH, Chater N (2008) Language as shaped by the brain. Behav Brain Sci 31:489-509

Cohen AH (1988) Evolution of the vertebrate central pattern generator for locomotion. In: Cohen A, Rossignol S, Grillner S (eds) Neural control of rhythmic movements. John Wiley & Sons, New York, NY, pp 129-166

Darwin C (1871) *The descent of man and selection in relation to sex*. John Murray, London

Davis, BL, MacNeilage PF (1995) The articulatory basis of babbling. J Speech Hear Res 38:1199-1211

Davis BL, MacNeilage PF, Matyear CL (2002) Acquisition of serial complexity in speech production: a comparison of phonetic and phonological approaches. Phonetica 59:75-107

Davis JW, Richards WA (2000) Relating categories of animal motion. Ohio State University, Dept. of Communication and Information Science Technical Report OSU-CISRC-11/00-TR 25

Dawkins R (1976) The selfish gene. Oxford University Press, Oxford

Dolata J, Davis BL, MacNeilage PF (2008) Characteristics of the rhythmic organization of babbling: implications for an amodal linguistic rhythm. Infant Behav Dev 31:422-431

Donald M (1991) Origins of the modern mind. Harvard University Press, Cambridge, MA

Dunbar RIM (1996) Grooming, gossip and the evolution of language. Harvard University Press, Cambridge, MA

Eibl-Eibesfeldt I (1989) Human ethology. Aldine de Gruyter, New York, NY

Elman JL, Bates EA, Johnson MH, Karmiloff-Smith A, Parisi D, Plunkett K (1996) Rethinking innateness: a connectionist perspective on development. MIT Press, Cambridge, MA

Evans N, Levinson SC (2009) The myth of language universals: language diversity and its importance for cognitive science. Behav Brain Sci 32:429-448

Fentress JC, Gadbois S (2001) The development of action sequences. In: Blass, E (ed) Handbook of behavioral neurobiology, Vol. 13. Kluwer Academic/Plenum Publishers, New York, NY, pp 393-431

Fitch WT (in press) Innateness in language: a biological perspective. In: Tallerman M, Gibson KR (eds) Encyclopedia of language evolution. Oxford University Press, Oxford

Gazzaniga MS, Ivry R, Mangun G (1999) Cognitive neuroscience: the biology of the mind. MIT Press, Cambridge, MA

Giraud A-L, Kleinschmidt A, Poeppel D, Lund TE, Frackowiak RSJ, Laufs H (2007) Endogenous cortical rhythms determine cerebral specialization for speech production and perception. Neuron 56:1127-1134

Goodman CS, Coughlin BC (2000) Introduction: the evolution of evo-devo biology. Proc Natl Acad Sci 97:424-425

Hiiemae KM, Palmer JB (2003) Tongue movements in feeding and speech. Crit Rev Oral Biol M 14:413-429

Hogan JA (2001) Development of behavioral systems. In: Blass E (ed) Handbook of behavioral neurobiology, Vol. 13. Kluwer Academic/Plenum Publishers, New York NY, pp 229-279

Hohenberger A, Happ D, Leuninger H (2002) Modality-dependent aspects of sign language production: evidence from slips of the hands and their repairs in German sign language. In: Meier RP, Cormier K, Quinto-Pozos D (eds) Modality and structure in signed and spoken languages. Cambridge University Press, Cambridge, pp 112-142

Houghton G (1990) The problem of serial order: a neural network model of sentence learning and recall. In: Dale R, Mellish C, Zock M (eds) Current research in natural language generation, Academic Press, London, pp 287-319

Houghton G, Hartley T (1995) Parallel models of serial behavior: Lashley revisited. Psyche 2:2-25

Jakobson R (1967) About the relation between visual and auditory signs. In: Wathen-Dunn W (ed) Models for the perception of speech and visual form. MIT Press, Cambridge, MA, pp 1-7

Jakobson R (1941/1968) Child language, aphasia, and phonological universals. The Hague, Mouton

Jenkins JJ (2010) Book review: MacNeilage's the origin of speech. Cog Crit 2:141-150

Kern S, Davis BL (2009) Emergent complexity in early vocal acquisition: cross-linguistic comparisons of canonical babbling. In: Chitoran I, Coupé C, Marsico E, Pellegrino F (eds) Approaches to phonological complexity. Mouton de Gruyter, Berlin, pp 353-376

Ladefoged P (1993) A course in phonetics, 5th edn. Harcourt Brace Jovanovich, New York, NY

Ladefoged P (2005) Features and parameters for different purposes. IN: Working Papers in Phonetics 104, Department of Linguistics, UCLA. Los Angeles, Retrieved from: http://escholarship.org/uc/item/3sf437w9

Lashley KS (1951) The problem of serial order in behavior. In: Jeffress LA (ed) Cerebral mechanisms in behavior: the Hixon symposium. Wiley, New York, NY, pp 112-136

Levelt WJM (1992) Accessing words in speech production: stages, processes and representations. Cognition 48:1-22

Locke JL (1983) Phonological acquisition and change. Academic Press, New York, NY

Lund JP Kolta A (2006) Brainstem circuits that control mastication: do they have anything to say during speech? J Commun Disord 39: 381-390

MacNeilage PF (1997) Acquisition of speech. In: Hardcastle WJ, Laver J (eds) Handbook of phonetic sciences. Blackwell Publishing, Oxford, pp 301-322

MacNeilage PF (1998) The frame/content theory of evolution of speech production. Behav Brain Sci 21: 499-548

MacNeilage PF (1999) Whatever happened to articulate speech? In: Corballis MC, Lea SEG (eds) The descent of mind: psychological perspectives on hominid evolution. Oxford University Press, Oxford, pp 116-137

MacNeilage PF (2008a) The origin of speech. Oxford University Press, Oxford

MacNeilage PF (2008b) The frame/content theory. In: Davis BL, Zajdó K (eds) The syllable in speech production. Lawrence Erlbaum Associates, Taylor & Francis Group, New York, NY, pp 1-28

MacNeilage PF (in press, a) Sound patterns of first words and how they became linked with concepts. In: Lefebvre C (ed) The evolution of language. Cambridge University Press,Cambridge

MacNeilage PF (In press, b) Lashley's serial order problem and the acquisition of learnable vocal and manual communication. In: Vilain A (ed) Primate vocalisation and human language: vocalisation, imitation, and deixis in humans and non-humans. Oxford University Press, Oxford

MacNeilage PF, Davis BL (1990) Acquisition of speech production: frames, then content. In: Jeannerod M (ed) Attention and performance X111: motor representation and control. Erlbaum, Hillsdale, NJ, pp 453-476

MacNeilage PF, Davis BL (2000) On the origin of the internal structure of word forms. Science 288: 527-531

MacNeilage PF, Davis BL (2001) Motor mechanisms in speech ontogeny: Phylogenetic, neurobiological and linguistic implications. Curr Opin Neurobiol 11: 696-700

MacNeilage PF, Davis BL (2002) On the origins of intersyllabic complexity. In: Givón T, Malle BF (eds) The evolution of language out of prelanguage. J.V. Benjamins, Amsterdam, pp 155-170

MacNeilage PF, Davis BL, Kinney A, Matyear CL (1999) Origin of serial output complexity in speech. Psychol Sci 10:459-460

MacNeilage PF, Davis BL, Kinney A, Matyear CL (2000) The motor core of speech: a comparison of serial organization patterns in infants and languages. Child Dev 7:153-163

Maddieson I (1999) In search of universals. Proceedings of the 14th International Congress of Phonetic Sciences. University of California Press, 3:2521-2528

Meier RP, Willerman R (1995) Prelinguistic gesture in deaf and hearing infants. In: Emmorey K, Reilly J (eds) Language, gesture and space. Erlbaum, Hillsdale, NJ, pp 391-409

Mielke J (2008) The emergence of distinctive features. Oxford University Press, Oxford

Miller GA (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol Rev 63: 343-355

Mitchell PR, Kent RD (1990) Phonetic variation in multisyllable babbling. J Child Lang 17:247-265

Oller DK (1980) The emergence of speech sounds in infancy. In: Yeni-Komshian G, Kavanagh JF, Ferguson GA (eds) Child phonology 1. production. Academic Press, New York, NY, pp 93-112

Oller DK (2000) The emergence of the speech capacity. Erlbaum, Mahwah, NJ

Petitto LA, Marentette PF (1991) Babbling in the manual mode: evidence for the ontogeny of language. Science 251:1493-1496

Pinker S (1994) The language instinct. William Morrow, New York, NY

Prince A, Smolensky P (1997) Optimality: from neural networks to universal grammar. Science, 275:1604-1610

Ridley M (2003) Nature via nurture: genes, experience and what makes us human. Harper Collins, New York, NY

Ronnqvist L, von Hofsten C (1994) Varieties and determinants of finger movements in neonates. Early Dev Parenting 3:81-94

Rousset I (2003) From lexical to syllabic organization: favored and disfavored co-occurrences. Proceedings of the 15th International Congress of Phonetics. Autonomous University of Barcelona, Barcelona, pp 2705-2708

Shattuck-Hufnagel S (1979) Speech errors as evidence for a serial ordering mechanism in speech production. In: Cooper WE, Walker ECT (eds) Sentence processing: psycholinguistic studies presented to Merrill Garrett. Erlbaum, Hillsdale, NJ, pp 295-342

Shubin N, Tabin C, Carroll S (2009) Deep homology and the origins of evolutionary novelty. Nature 45:818-823

Smith BL, Brown-Sweeney S, Stoel-Gammon C (1989) A quantitative analysis of reduplicated and variegated babbling. First Language 9:175-189

Thelen E (1981) Rhythmical behavior in infants: an ethological perspective. Dev Psychol 17: 237-257

Tooby J, Cosmides L (1992) The psychological foundations of culture. In: Barkow JH, Cosmides L, Tooby J (eds) The adapted mind: evolutionary psychology and the generation of culture. Oxford University Press, New York, NY, pp 19-136

van der Stelt JM, Koopmans-van Beinum FJ (1986) Early stages in the development of speech movements. In: Lindblom B, Zetterstrom R (eds) Precursors of early speech. Stockton Press, New York, NY, pp 37-50

Van Hooff, JARAM (1967) Facial displays of the catarrhine monkeys and apes. In: Morris D (ed) Primate ethology. Weidenfield and Nicholson, London, pp 7-68

Vousden JI, Brown GDA, Harley TA (2000) Serial control of phonology in speech production: a hierarchical model. Cognitive Sci 41:101-175