Particulate Speech:

The Emergence of the Phoneme from Syllable Frame Structures

Peter F. MacNeilage

University of Texas at Austin

macneilage@psy.utexas.edu

Abstract

William Abler (1989) has postulated the "particulate principle of self-diversifying systems … which holds that when systems form structures having an infinite range of properties, such systems must be based on particles …". According to Abler, natural systems exhibiting this principle "include chemical interaction, biological inheritance and human language". The present focus will be on the linguistic particles singled out by Abler—the phonemes of language. According to Abler's physics-based conception, the three systems "derive their properties by common inheritance" from "some property of material nature". However, in the present view physics and biology differ in that the source of ultimate causes in biology is the theory of evolution by natural selection, and phonemes must be explained in terms of this theory, not in terms of the laws of physics. As to the origin of the three instances of particulation, the particles of chemical interaction, and biological inheritance—elements and nucleic acids respectively—were synthesized from pre-existing subcomponents, while phonemes may have emerged from a pre-existing syllable frame structure, according to the author's "Frame/Content" theory of evolution of speech production, reviewed here. Two subphonemic particles of speech have also been proposed—the "distinctive feature" of structural linguistics and the "gesture" of proponents of dynamical systems theory. However, these particles do not exist in nature. In conclusion, syllable frame structures can be considered to provide a unique pre-systematic basis for both the evolution and the development of phonemic systems of languages.

Introduction

William Abler has postulated the *"particulate principle of self-diversifying systems ... which holds that when systems form structures having an infinite range of properties, such systems must be based on particles ...."* (Abler, 1989, p. 1). According to Abler, *"These systems include (1) the chemical elements, whose underlying particulate units are atoms, (2) biological inheritance, whose particulate units are genes, and (3) human language, whose underlying particulate units are phonemes and morphemes" (p. 1)*. Abler designates the three systems as *"Humboldt Systems"* in that they *"share the further property of conforming to Humboldt's (1836) characterization of language, i.e., all systems 'make infinite use of finite media' (p. 70) whose synthesis creates something that is not present per se in any of the associated constituents' (p. 67)" (p. 1)*.

The present author concurs with Abler's view that the relevant particle in linguistics is the phoneme. A phoneme is a sound—a consonant or vowel—that differentiates one word from another. For example, the words "ten" and "den are different by virtue of their initial consonants, denoted in phonemic transcription as /t/ and /d/. Self-diversification in this context refers to the way in which segments are concatenated in many ways to form the meaning units, or morphemes of words. According to the author's Frame/Content Theory (MacNeilage, 1998; 2008) phonemes (collectively content) emerge, in both evolution and development from a superordinate syllable frame structure. This theory, and the evidence for it will be presented, before considering two other putative speech units, the "distinctive feature" and the "gesture", and before considering Abler's proposed explanation for the existence of the particulate principle.

The Frame/Content Theory

Phylogeny

Speech can be characterized as a voice source (described as phonation or voicing) modulated by a series of mouth closing/opening alternations. An alternation cycle is a syllable, and the closing and opening phases are consonants and vowels, respectively. In an important paper entitled "The Problem of Serial Order in Behavior" Lashley (1951) noted that evidence regarding the serial organization of speech comes from serial-ordering errors in normal speakers because these errors not only reveal the units of speech—phonemes, consisting of consonants and vowels—but also the constraints on how they are ordered. In these errors (e.g., spoonerisms such as "catnip" → "natkip" and "ad hoc" → "odd hack"), misplaced consonant and vowel forms (segments) end up in the same position in syllable structure that they originated in. The widespread nature of this syllable structure constraint on segmental positioning led Levelt, the most prominent figure in the psycholinguistics of language production, 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" (1992, p. 10; italics mine). The Frame/Content theory is a theory about the origin of this form of organization, in both phylogeny and ontogeny.

One might ask, what is so remarkable about this form of organization. Well, for one thing, it doesn't happen in typing, a linguistic serial-ordering operation that we invented. Here, we are just as likely to reverse consonant and vowel letters as we are to reverse pairs of consonant letters, or pairs of vowel letters (MacNeilage, 1985). So the fact that, for speech, nature has chosen something different from a simple linear ordering of units (as in keystrokes) that occasionally suffers from little misalignments of these units seems to be an important clue as to how we came to be able to do speech.

According to the Frame/Content theory, the syllable frame evolved from cycles of mouth close-open alternation (closed for consonants, open for vowels) controlled by the mandible. A cyclicity of this general form evolved in mammals for acts of ingestion (chewing, sucking, and licking) about 200 million years ago, and we probably exapted (borrowed) it for speech. As the two movements in the cycle are antagonistic—they go in opposite directions in successive phases of the cycle—and as the two major-

class segmental categories (consonants and vowels) probably evolved from these two opposite movements, there was never an opportunity for consonants and vowels to get mixed up with each other in the control of the cycle, hence the syllable structure constraint on speech errors. Following the evolution of ingestive cyclicities, there may have been an intermediate stage of use of this cycle for visuofacial communicative cyclicities (e.g., lipsmacks, tonguesmacks, teeth chatters). These communicative sequences of close/open alternations are common in modern monkeys (Redican, 1975). In the step from smacks to speech, the cycles of mandibular oscillation were paired with vocal-fold vibration (voicing, phonation) to form protosyllable sequences such as "bababa." The suggestion that protosyllables like this may have been the original phonological forms of speech is endorsed in the recent influential monograph by Jackendoff (2002).

How was this putative sequence of events chosen by natural selection. The advent of internal temperature regulation in the transition from reptiles to mammals required higher rates of food ingestion, and the original mandibular cyclicities for chewing, sucking, and licking can all be regarded as increments in food-processing efficiency selected for that end. Van Hooff (1967) has suggested that smacks may have evolved their communicative status from cyclical ingestive movements elicited in a manual-grooming context. Animals looking forward to finding a food item such as a salt grain in individual instances of grooming might have begun chewing movements in anticipation of such a discovery. Initial selection pressures for the protosyllabic cyclical forms (in effect, smacks with voicing) may have come from vocal grooming, whereby, according to Dunbar (1996), vocalization may have been substituted for manual tactile contact as ancestral hominid group sizes increased enough to make the latter behaviour ineffective as a device for social cohesion. (Both manual and vocal grooming are widely regarded as having beneficial consequences for social organization.) It's also possible that an evolving capacity to learn vocalization—a capacity not present in other living primates—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, almost as salient and unique in humans as is speech, is evident in modern human music, dance, opera, ballet, movies, games, sports, etc. However, there has been a marked reluctance to give Donald's claim of a fundamental status for this capability the respect it deserves in the domain of human phylogeny.

Ontogeny

According to the frame/content theory, the process of speech acquisition may throw light on its evolution. A compelling consideration here is how rhythmic infant babbling is from its very beginning, a basic action available to all normal infants. There is no sign of a process whereby infants gradually make their output more rhythmic as there would be if they were learning this capacity, in the way, for example, that one learns to make a smooth, rhythmic golf swing. For example, we have found that 2 successive close-open cycles of babbling, which average 1/3 of a second each, differ from each other, on the average, by only 1/30th of a second (Dolata, Davis and MacNeilage, 2008). The degree of precision in this cyclical action of a 7 month old infant, from its very beginning, even though it is not a life supporting action, unlike the heart beat or respiration, suggests that the underlying function has an extremely fundamental status. Furthermore, as we will shortly see, the biomechanics of this early speech-like output of young modern hominids are extremely simple, suggesting that the biomechanics of the original ancestral speech-like actions must have been quite simple too, and, of course, similar. This, among other considerations points to the conclusion that babbling has evolved as an innate substructure for speech (MacNeilage, in press).

Over a generation of work on speech acquisition with Barbara Davis and others (see, for example, MacNeilage and Davis, 2000) has led us to the conclusion that the initial babbling stage of speech development, in which the same syllable tends to be "reduplicated" (as in "bababa"), can be regarded as a "Frame" stage, with little in the way of internal "Content." We have found, and others have confirmed, that the production of both babbling and early words is mainly a matter of oscillation of the mandible,

giving rise to an alternation between a closed- and open-mouth configuration. In a number of studies we have shown that sequential adjustments of tongue position, which is the main determinant of the different consonants and vowels of a language, tend not to occur in early babbling and speech episodes (See Davis and MacNeilage, 2002 and MacNeilage and Davis, 2002 for summaries).

This is a sufficiently central property of speech acquisition to deserve some detailed illustration. For example, if the tongue is in the front of the mouth when an infant begins a babbling episode or a word, it tends to stay there throughout. Consequently, tongue-front consonants such as [t], [d] and [n] tend to accompanied by front vowels in consonant-vowel (CV) syllables. Examples of words with these form are "Ted," "dead," and "Ned." A similar relation is found between tongue-back consonants ([k] and [g]) and back vowels, as in words such as "cook" and "gook." Perhaps most interestingly, the vowels produced following lip consonants (e.g., [p], [b], [m]), such as the first vowel in "papa" and the vowel in "bubble," and the so-called neutral vowel in the first word of "the box" are called "central vowels" and are probably made with the tongue in its resting position in the center of the mouth. In all these cases, the consonant-vowel sequence, even when it's repeated a number of times, does not involve active tongue movement at all. We have called this overall phenomenon "Frame Dominance" (Davis and MacNeilage, 1995, p. 1199).

Thus, apart from the cyclical frame, the operation of the speech apparatus in babbling and early words is characterized primarily by biomechanical inertia. The tongue is apparently not called upon to change its position during a vocal episode. Davis and I reasoned that if these inertial patterns were also favoured in modern languages, this would constitute evidence for their fundamental status in the evolution of speech. In our study of 10 languages (MacNeilage et al, 2000) and in a study of 14 languages by Rousset (2003), the three basic patterns described above were present in about 70% of instances, compared with an expected base rate of 33%

There is one other basic relation between babbling and speech patterns of language that suggests the former is fundamental to the latter. It is that the consonant-vowel (CV) form favoured in these infant vocalizations is, with a single exception, (Breen and Pensalfini, 1999) the only universal syllable type in the world's languages.

If indeed the first stage in the evolution of speech was a frame stage, how then did we eventually evolve the capacity to program individual segments into syllable frame structures using a frame/content mode, thus achieving "segmental independence (MacNeilage and Davis 1990b)? A clue as to how this might have begun to happen comes from the course of events in speech acquisition. There is a strong trend across language environments for the first systematic step toward intersyllabic variegation 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." In studies of adult languages, my colleagues and I, (MacNeilage et al., 2000) as well as Rousset (2003), find that, as in the case of the tongue-inertia tendencies, this tendency is also rife in languages. Davis and I have suggested that this is a result of self-organization (MacNeilage and Davis, 2000). It may be a case of beginning simply with what we call a "Pure Frame," (a lip consonant-central vowel sequence produced only by the mandible) because, in lay terms, it is easier. It doesn't also involve beginning with a non-resting tongue position. 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 supported by evidence from neurophysiology that starting to move is a special problem for the brain, because it is addressed with dedicated circuitry (e.g. Gazzaniga and Heatherton, 2003).

At a more general level, the emergence of the segments comprising content from syllable frames during the course of early development can be viewed in the following way. Consider first the case of pure frames in which no independent tongue action accompanies the mandibular cycle for the CV syllable. Resulting in a labial (lip) consonant and a central tongue position for the vowel. What is necessary, in order to produce this kind of consonant and vowel in adult speech along with other segments, at the rate

of about 15 segments per second is that we develop the capacity to actively control lip closure for the consonant and tongue control for the assumption of the required central tongue position. Electromyographic studies of muscle action potentials during adult speech show the presence of this active command structure (e.g. MacNeilage and DeClerk, 1969). Obviously the commands to the muscles must be restricted to the time windows allowed for the movement associated with the labial consonant and the movement associate with the central vowel at this speaking rate. Once this active control of individual labial consonants and central vowels has been achieved, these segments can be produced in contexts other than the one in which they co-occur with each other, as passive resultants of mandibular oscillation.

Consider now the development of control of individual content segments from the beginning point of the favoured frames that involve the tongue (coronal-front and dorsal-back CV forms). In this case, when the lingual consonants and their preferentially associated vowels are not called for together in target words, it is necessary to learn to restrict the tongue positions associated with the individual segments to single consonants (coronal or dorsal) and single vowels (front or back).

In summary, at the level of peripheral control, the ability to program individual segments in the case of pure frames requires generation of segment-specific movement control capacities where none existed before. In contrast differentiation of segments from lingual frames requires the restriction of the tongue positioning movement to the time window of a single segment, be it the consonant or the vowel. In the particular case in which the LC sequence is in the word being attempted, this ideally means restricting the tongue movement to the final consonant if the vowel is not a front vowel, though one might expect, at early stages that there might be some tendency for the vowel preceding the final consonant to share its front place of articulation because it might take some time to reliably make the movement specific to the coronal consonant. (See Lindblom and MacNeilage, 2011, for further discussion.)

Davis and I have not pursued the specifics of what we have called the "Frames, then Content" progression (MacNeilage and Davis, 1990a) of the infant/child in speech acquisition beyond the identification of the labial-coronal effect. The obvious question to ask is whether, as in the case of adults, serial ordering errors of infants throw light on this progression. Fortunately, we have available a large scale study of speech errors during speech acquisition. Jaeger (2005) has analyzed 1,383 speech errors made by 3 infants/children within the age range of 1-5 years. In a summary of her findings she states that "An extremely valuable heuristic in which to frame this discussion is the 'Frames, then Content' developmental progression proposed by MacNeilage and Davis (1990a) …" (p. 214). Her conclusion from an analysis of these errors is that "The discussion … regarding the developmental pathways taken by children towards the learning of phonology is completely consistent with MacNeilage and Davis's (1990a) model of language acquisition." (p. 219).

At this point it is necessary to address the question of whether phonemes are indeed discrete, as Abler claims, in a way that the elements of chemistry and the chemical bases of inheritance are. The claim that phonemes are discrete is substantiated by speech errors that show that single segments are independently controllable in otherwise correct sequences. In addition to exchange errors there are 4 other types of errors in which segments act as independent entities (Shattuck-Hufnagel, 1979; MacNeilage, 2008, pp. 81-83); Shift errors: A unit disappears from its correct location and appears in another location—e.g. State loaned and (owned land): Addition errors: an extra unit is added to an intended sequence—e.g. spublic speaking (public): Omission errors: a unit is dropped from an intended sequence—e.g. sonata umber ten (number): Substitution errors: One intended unit is replaced by another—e.g. it's a shallower test (chest).

Evidence of a Relation Between Lipsmacks and Babbling.

A key evolutionary hypothesis of the F/C theory is that syllables evolved from lipsmacks. Important behavioural evidence in support of this claim has recently been provided by Morrill et al., (in press) in a

paper entitled "Monkey Lipsmacking Develops Like Speech". In a developmental study of the dynamics of lipsmacks in macaque monkeys they explicitly tested the F/C hypothesis that syllables evolved from lipsmacks. They report that neonatal monkeys produce lipsmacks at the rate of about 3 per second, a rate similar to that of syllable production in infant babbling (Dolata, Davis and MacNeilage, 2008), while adult monkeys produced lipsmacks at the rate of about 5 per second, a rate similar to that of syllable production in adult speech (Dolata, Davis and MacNeilage, 2008). They also found that the chewing rate in these animals was initially slower than the neonatal lipsmack rate (about 2 per second), and this chewing rate did not change very much with development. The similarity of these repetition rates for neonatal and adult lipsmacks on the one hand and babbling and speech on the other is highly suggestive evidence that speech could have evolved from lipsmacks.

Comparative Neurobiological Evidence for the Syllable Frame.

It is well known in the comparative neurobiology of action that the production of biphasic cycles in general is controlled by neurophysiological entities known as a "Central Pattern Generators" (CPGs) A CPG is defined as "any network within the CNS that coordinates a motor behaviour or part thereof" (2006, p. 751). Grillner notes that "We know in all animals, vertebrates or invertebrates, movements are controlled by CPG networks that determine appropriate sequences of muscle activation. Each animal is endowed with a broad repertoire of CPGs located in different regions of the central nervous system, and available for differential activation, thus providing animals with a distinctive set of solutions to accommodate their widely divergent patterns of behaviour" (2006, p. 751).

There are two main categories of non-communicative voluntary action in vertebrates, one related to locomotion, and the other to food attainment and ingestion. With regard to locomotion, CPGs generate biphasic oscillatory cycles (e.g., for walking, swimming, and flying). Georgopoulos and Grillner (1989) point out that with the advent of terrestrial mammals, an increasing ability to control exact foot preference is superimposed on basic locomotory cycles. They also propose that in primates "the precise forelimb movements used to position the limb at will and to grasp different objects are very similar to, and have evolved from, those used to position the limb accurately during locomotion" (p. 1210).

With respect to ingestion, one can envisage a similar progression from a basic cycle to a modulation of that cycle. As part of selection for increasing food processing and ingestion in newly evolving warm-blooded mammals about 200 million years ago, an oscillatory CPG controlling the mandible for chewing, sucking, and licking may have developed, and in omnivorous forms ancestral to humans, an increasing ability to modulate the biphasic cycle for special processing purposes, especially those related to chewing, has followed. Most specifically, Lund and Kolta (2006) have proposed that brainstem circuits associated with the CPG that controls mastication "also participate in the control of human speech" (p. 381).

Historically, the main brain region associated with the motor control of speech is an area of the posterior inferior frontal lobe encompassing Broca's area and its immediate surround. This area includes sites associated with the control of mastication in monkeys. Furthermore mirror neurons associated with lipsmacking have been found in area F5 of monkeys, which is considered to be the homolog of Broca's Area in monkeys. Until recently, F5 had not been thought to be significantly involved in monkey vocalization. However, Coude et al. (2011) have recently recorded neurons in monkey F5 which discharged prior to and/or concurrent with the production of conditioned coo vocalizations in two monkeys.

Evidence for a specific role of syllable control in the posterior inferior frontal cortex of humans has been presented by Giraud et al. (2007). They 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 substrates" (p. 1133).

In addition to this possible involvement of posterior inferior frontal cortex, I have argued that with the evolution of speech from prespeech communicative capabilities, aspects of superordinate control of the mandibular cycle, or motor frame for speech moved from this region of frontal cortex to posterior medial frontal cortex. This control became mediated by 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 posterior medial cortex as a site for frame generation.

This medial region was presumably responsible for the rhythmic repetitions of the same CV in Broca's famous patient "Tan", and similar repetitive 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.

Finally, in recent years, comparative neurobiological evidence suggesting an evolutionary relation between lipsmacks and syllable frames comes from studies of the anterior insula. The insula is a region of cerebral cortex lying deep within the lateral sulcus, (sylvian fissure) intermediate between the superior medial surface and the inferior medial surfaces of the sulcus. The initial indication that the anterior insula might be involved in the evolution of speech involved a study of "Apraxia of Speech". Apraxia of speech (AOS) is "a syndrome assumed to reflect impaired higher order aspects of speech motor control, i.e. the "planning" of the respective vocal tract movement sequences" (Ackerman and Riecker, 2010, p 420). In a surprising finding in 1996, Dronkers reported that the only area of brain damage common to a group of N patients with apraxia of speech was the anterior region of the insula in the dominant hemisphere. Ackerman and Rieker (2010) summarize subsequent imaging studies that confirm a role of the anterior insula in the planning of speech production. They conclude that the region is "sensitive to phonetic-linguistic structure of verbal utterances" (p. 419), particularly emphasizing a study in which haemodynamic activity in the region is sensitive to the complexity of syllable structure.

Ackerman and Rieker observed that "considering … the well-established participation of the anterior insula in the phylogenetically rather old perceptual-motor functions such as feeding-related behavior … it is a bit bewildering that 'motor planning' of speech (Dronkers, 1996) or 'motor articulatory planning' (Nestor, 2003), capacities unique to our species, might might depend on the rostral, i.e. 'paralimbic' component of intrasylvian cortex." (p. 420) However, rather than being bewildering, it is what would be expected if one believes that the articulatory capabilities of speech evolved from food ingestion capabilities. Numerous imaging studies have shown involvement of the insula in both swallowing and chewing. The specific involvement of the insula in the motor action of human chewing, as distinct from effects of gustatory and olfactory input related to the chewed object, is indicated by a study in which electrical stimulation of the insula of epileptic patients produced chewing movements (Ostrovsky et al, 2000).

Interestingly, in the present context, these authors also noted the elicitation of "lipsmacks" in one patient. These might have been instances of pure frames unaccompanied by phonation. Thus, as might have been expected from the F/C perspective, human frames or framelike actions may have had a phylogenetic precursor in insular cortex. In a further confirmation of a phylogenetic relation between lipsmacks and syllables, Caruana et al. (2011) have elicited lipsmacks in monkeys by stimulation of the insula. These

lipsmacks were dependent on social context, occurring only when the monkey was face to face with another individual, a circumstance that could presumably have been carried over into humans. The fundamental role of lipsmacks in monkey communication is emphasized by the finding of Ferrari et al (2009) that neonatal monkeys exchange lipsmacks with their mothers at birth. It may also be of interest, in the light of the possibility that lipsmacks derived from chewing, that the lipsmacks evoked by Caruana et al. sometimes terminated in chewing movements.

Subphonemic Particles?

Are phonemes the minimal particles of speech, or are they composed of smaller units? In the view of Studdert-Kennedy (1998), "consonants and vowels are not primitive units, but compounds, analogs of the molecule not the atom" (p. 209). There have been two candidates for the supposed subphonemic units of speech, the "distinctive feature" and the "gesture". A "distinctive feature" is an attribute of a segment that distinguishes the segment from another. For example the word "ten" is distinguished from the word "den" by a distinctive feature of voicing. The distinctive features are considered by the modern discipline of generative phonology to constitute a finite set of innate abstract entities out of which sound systems of languages are made. (See MacNeilage, 2008, pp. 225-235 for discussion.) Phonemes are considered to be bundles of distinctive features. But as Studdert-Kennedy (1998) points out, features "have no substance" (p. 210). As is evident from the adjectival terminology of all feature theories (nasal, coronal, continuant etc.) features are not independent entities (like atoms) but properties or attributes of the segments they describe. They are not action components. Instead, as Ladefoged (2006) points out, they "are best regarded as artifacts that linguists have devised in order to describe linguistic systems" (p. 12).

An alternative subsegmental action unit called the "gesture" has been proposed by members of a subdiscipline derived from dynamical systems theory called Articulatory Phonology, (e.g. Goldstein and Fowler, 2003) and endorsed by Studdert-Kennedy (1998; 2000). These gestures are actions involving 6 putative basic particles or "atoms" of speech, also called "organs". These particles/atoms, supposedly associated with 6 speech "organs" are the lips, the tongue tip, the tongue body, the tongue root, the soft palate and the glottis (Goldstein, Byrd & Saltzman, 2006). The gestures are considered to be oscillators, each involving the making and releasing of a vocal tract constriction. The "atoms" are considered to be combined into "molecules" (consonantal and vocalic forms) by naturally occurring contingencies arising from the coupling of oscillators, and influenced in their serial organization by a dimension of articulatory compatibility/incompatibility between pairs of gestures. (Goldstein, Byrd & Saltzman, p. 217)

The main problem with this conception is that, as we will now see, the use of the metaphors of "atom" and "molecule" is quite spurious. Atoms are the basic particles of matter and they combine into molecules to form matter. There are 118 atoms (called elements) of which 98 occur naturally, and they can be arranged consecutively in the Periodic Table according to the atomic weights indicated by their numbers (1 to 118). The individual structures are known in detail and precise rules for the way they can combine into molecules are also known (see Wikipedia—"Periodic Table" for discussion).

Nothing remotely resembling the state of affairs regarding atoms and molecules is true of the subsegmental/segmental level of speech production. The closest one can come to an analogy between chemical units and articulators is in the case of the soft palate, as it can function independently of other articulators. It typically makes a single binary contribution to speech (nasal, non-nasal) in a manner similar to the way that an element might participate or not participate in a molecule. But it clearly is different from the elements of chemistry because it retains its quality when combined, in a way that elements (e.g. oxygen in water) do not. Beyond that, the lower component of the lips, and the tongue are functionally dependent on the mandible, and are therefore not independently definable from an articulatory standpoint. Most importantly, there is no justification for regarding the three regions of the tongue (tip, body and root) as separate organs, across the board. The common sense conclusion that the tongue is not three organs but one is the correct conclusion. The main manifestation of its unity as an

organ is the fact that one of its muscles, the genioglossus muscle, originating at the chin, fans radially into the entire tongue. Although it is not necessarily always activated as a whole, there is no evidence that it has 3 separate functional subdivisions. Why are there not 4 organs? Why is the tongue blade, often a separate component in descriptions of the tongue, not also included as one of the "atoms" of speech? And why is the mandible not considered to be an "atom" of speech, given that it is clearly a discrete organ anatomically, and it is almost as important in the formation of constrictions, and their releases, as the tongue is? In short, with the exception of the soft palate, where discreteness is self evident, discreteness of the putative subsegmental gestures has not been established.

For a gestural perspective to be successful, it should be able to characterize the different kinds of consonants and vowels in languages. But in contrast to the realm of generative phonology, in which the unit "distinctive feature" is roughly co-extensive with the gesture, (see MacNeilage, 2008 for discussion) such a comprehensive characterization has never been attempted. A fundamental issue here is that finiteness entails discreteness. One can't count entities unless they are discrete, and, with the exception of the soft palate, the discreteness of gestures is not evident. But beyond this, consider the problem at an empirical level: in a count of the segmental content of 317 languages Maddieson (1994) found a total of 558 consonants and 210 vowels. The main contributor to this huge number of segments is tongue position. How might these individual sounds be constructed from the 6 "atoms" of articulatory phonology, including only 3 "atoms" involving the tongue? Not only is no attempt made to address this problem, but conception of the tongue as three organs is not even used in the approaches of the AP group to either phylogeny or ontogeny, beyond the assertion that these organs exist.

Explaining the Particulate Principle

How should the common manifestation of the particulate principle in the domains of chemical interaction, inheritance and language be explained. According to Abler "the several particulate systems derive their properties by common inheritance from …. the particulate principle of self-diversifying systems which predates them all." (p. 1) Abler believes that the principle itself derives from "some underlying property of material nature." (p. 12). By material nature he means some property of the physical world. This property is considered to exist a priori, as in the "essences" of Plato. (See MacNeilage, 2008 for a discussion of essentialism with reference to language in particular.). "Particulation" is given axiomatic status as "a Euclidean principle which acts in a universe governed by classical arithmetic" (p. 12). As to the origin of the phoneme in particular, it is considered to represent "a class of discrete mental categories which probably depends, at some level, on the discrete, or particulate, nerve impulse…." (p 12).

It can be seen that Abler has a reductionistic view of physics as the basis of lawfulness in other domains.

In spite of the conviction that all sciences ought to be based on the same principles because all are based on physics, scientists use different systems of rules in dealing with various natural systems, implicitly treating such systems as if they were unrelated. The discrepancies are nowhere greater than between social, biological and physical sciences.

It is obvious that the present perspective is somewhat different from Abler's perspective. In general terms, it derives from Crick's contention that the mechanism of natural selection "makes biology different from all the other sciences." (Crick, 1988, p. 5) As the phoneme evolved by natural selection, it cannot be explained by the laws of physics. Instead, it evolved via a typical mode of selection—it evolved by tinkering (Jacob, 1977). First, the central pattern generator for ingestion was tinkered into use for vocal communication in the form of the syllable frame. Then the two phases of the frame cycle were tinkered into use as programmable slots enabling the differentiation of linguistic messages by means of phonemes. This mode of origin is different in its overall form from the origin of the particulate principle in both the domain of chemical interaction and in the domain of inheritance. The phoneme emerged from a superordinate structure. Consideration of the origins of the chemical elements (Penzias, 1978) and the 4 chemical bases of inheritance, Freeland and Illardo (in press) shows that these particles were synthesized

from subcomponents— elements from protons and neutrons and electrons, and the nucleic acids constituting DNA and RNA, the two substances associated with the genetic code, from elements.

Thus, in the present view, the basis for the evolution and development of phonemic systems is a pre-systematic frame. In the case of the labial consonants and central vowels of pure frames, the emergence of the systematic basis of phonemes involves the provision of active control of articulatory positions which were previously produced passively, as inevitable side-effects of the two phases of the frame cycle. In the case of lingual frames the main task is to constrain tongue movements in the time domain to more or less coincide with the durations of the closing phase for consonants and the opening phase for vowels.

References

Abler, W. (1989). On the particulate principle of self-diversifying systems. *Journal of Social and Biological Structures, 12*: 1-13.

Ackermann, H. & Riecker, A. (2010). The contribution(s) of the insula to speech production: a review of the clinical and functional imaging literature. *Brain Structure and Function, 214*: 419-433.

Bohland, J.W., Bullock, S. & Guenther, F.H. (2010). Neural representations and mechanisms for the performance of simple speech sequences. *Journal of Cognitive Neuroscience, 22*: 1504-1529.

Breen, G. & Pensalfini, R. (1999) Arrernte: A language with no syllable onsets. *Linguistic Inquiry, 30*; 1-26.

Caruana F., Jazzini A., Sbriscia-Fioretti B., Rizzolatti G. & Gallese V (2011) Emotional and socialbehaviors elicited by electrical stimulation of the insula in the macaque monkey. *Current Biology 21*: 1-5.

Crick, F. (1988). *What Mad Pursuit: A Personal View of Scientific Discovery*. New York: Basic Books.

Davis, B. L.& MacNeilage, P. F. (1995). The articulatory basis of babbling. *Journal of Speech and Hearing Research, 38*: 1199-1211.

Davis, B. L. & MacNeilage, P. F. (2002). The internal structure of the syllable. In T. Givón & B. F. Malle (eds.), *The Evolution of Language out of Prelanguage*. Amsterdam: John Benjamins, 135-154.

Dolata, J., Davis, B.L. & MacNeilage, P.F. (2008). Characteristics of the rhythmic organization of babbling: Implications for an amodal linguistic rhythm. *Infant Behavior and Development 31*: pp. 422-431.

Donald, M. (1999). Preconditions for the evolution of protolanguages. In M.C. Corballis & S.E.G Lea (eds.), *The Descent of Mind: Psychological Perspectives on Hominid Evolution*. Oxford: Oxford University Press, 138-154.

Dronkers, N.F. (1996). A new brain region for coordinating speech articulation. Nature: 384, 159-161.

Dunbar, R.I.M. (1996). *Grooming, Gossip and the Evolution of Language*. Cambridge, MA: Harvard University Press.

Ferrari, P. F., Gallese, P., Rizzolatti, G. & Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth movements in the monkey ventral premotor cortex. *European Journal of Neuroscience, 17*: 1703-1714.

Ferrari, P.P., Paukner, A., Ionica, C. & Suomi, S.J. (2009). Reciprocal face-to-face communication between rhesus macaque mothers and their newborn infants. *Current Biology 19*: 1768-1772.

Freeland, S.J. & IIlardo, M. (In press). Language evolution in humans and ancient microbes: What can human language acquisition tell us about the origin of genetic information?

Gazzaniga, M. & Heatherton, T.F. (2003). *Psychological Science (2nd Edition)* New York: Norton.

Georgopoulos, A. P. & Grillner, S. (1989). Visual-motor coordination in reaching and locomotion. *Science, 245*: 1209-1210.

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

Goldstein, L.M. & Fowler, C.A. (2003). Articulatory phonology: A phonology for public language use. In N. Schiller & A. Meyer (eds). *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities.* Berlin: Mouton de Gruyter, 159-207.

Goldstein, Louis M., Byrd, D. & Saltzman,E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In Arbib, M. (ed.), *Action to Language Via the Mirror System.* Cambridge: Cambridge University Press, 215-249.

Grillner, S. (2006). Biological pattern generation: The cellular and computational logic of networks in motion. *Neuron, 52*: 751-766.

Humboldt, W. von (1836). Uber die Verscheidenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwichelung des Menschengeschlechts. Berlin: Royal Academy of Sciences. Republished (G.C. Buck and F. Raven, trans.) (1971) *Linguistic Variability and Intellectual Development.* Baltimore, Maryland: University of Miami Press.

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution.* Oxford: Oxford University Press.

Jacob, F. (1977). Evolution and tinkering. *Science, l96*: ll6l-1166.

Jaeger, J.J. (2005). *Kid's Slips: What Young Children's Slips of the Tongue Reveal About Language Development.* Mahwah, NJ: Lawrence Erlbaum Associates.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (ed.), *Cerebral Mechanisms in Behavior: The Hixon Symposium.* New York, NY: Wiley, 112-136.

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition, 48*, 1-22.

Lindblom, B. & MacNeilage, P.F. (2011). Coarticulation: A universal phonetic phenomenon with its roots in deep time. FONETIK 2011, 41-44 In – *QPSR TMH Vol 51*, KTH Stockholm, http://speech.kth.se/prod/publications/files/3588.pdf

Lund, J.P. & Kolta, A. (2006). Brainstem circuits that control mastication: Do they have anything to say during speech? *Journal of Communication Disorders, 39*: 381-390.

MacNeilage, P. F. (1985). Serial ordering errors in speech and typing. In V. A. Fromkin (Ed.), *Phonetic Linguistics.* New York, NY: Academic Press: 193-201.

MacNeilage, P. F. (1998). The Frame/Content theory of evolution of speech production. *Behavioral and Brain Sciences, 21*: 499-548.

MacNeilage, P.F. (2008). *The Origin of Speech.* Oxford: Oxford University Press.

MacNeilage, P.F. (In press). Sound Patterns of First Words and How They Became Linked With Concepts. In Lefebvre C (ed.) *The Origin of Language.* Cambridge University Press, Cambridge.

MacNeilage, P. F. & Davis, B. L. (1990a). Acquisition of speech production: Frames, then content. In M. Jeannerod (ed.), *Attention and Performance X111: Motor Representation and Control.* Hillsdale, N.J.: Erlbaum, 453-476.

MacNeilage, P.F. & Davis, B.L. (1990b). Acquisition of speech production: The achievement of segmental independence. In W.J. Hardcastle and A. Marchal (eds.), *Speech Production and Speech Modelling.* Dordrecht: Kluwer, 55-68.

MacNeilage, P. F. & Davis, B. L. (2000). On the origin of the internal structure of word forms. *Science, 288*: 527-531.

MacNeilage, P. F., & Davis, B. L. (2001). Motor mechanisms in speech ontogeny: Phylogenetic, neurobiological and linguistic implications. *Current Opinion in Neurobiology, 11*: 696-700.

MacNeilage, P. F. & Davis, B.L. (2002). The origins of intersyllabic complexity. In T. Givón & B. F. Malle (eds.), *The Evolution of Language out of Prelanguage.* Amsterdam: John Benjamins, 135-154.

MacNeilage, P.F. & De Clerk, J.L. On the motor control of coarticulation in CV syllables. *Journal of the Acoustical Society of America, l970: 45*, 1217-1233.

MacNeilage, P. F. Davis, B. L., Kinney, A. & Matyear, C. L. (1999). Origin of serial output complexity in speech. *Psychological Science, 10*: 459-460.

MacNeilage, P. F., Davis, B. L., Kinney, A. & Matyear, C. L. (2000). The motor core of speech: a comparison of serial organization patterns in infants and languages. *Child Development, 71*: 153-163.

Maddieson, I. (1984). *Patterns of Sounds.* Cambridge: Cambridge University Press.

Morrill, R.J., Paukner, A., Ferrari, P.F. & Ghazanfar, A.A. (In press). Monkey lip-smacking develops like human speech. *Developmental Science.*

Nestor, P.J., Graham, N.L., Fryer, T.D., Williams, G.B. Patterson, K. & Hodges, J.R. (2003). Progressive non-fluent aphasia is associated with hypometabolism centered on the left anterior insula. *Brain, 126*: 2406-2418.

Ostrowsky, K., Isnard, J., Ryvlin, P., Guenot, M., Fischer, C. & Mauguiere, F. (2000). Functional mapping of the insular cortex: clinical implication in temporal lobe epilepsy. *Epilepsia, 41*: 681-686.

Penzias, A.A. (1978). The origin of elements. (Nobel Lecture)
http://nobelprize.org/nobel_prizes/physics/laureates/1978/penzias-lecture.pdf

Redican, W. K. (1975). Facial expressions in nonhuman primates. In L. A. Rosenblum (ed.), *Primate Behavior: Developments in Field and Laboratory Research, Vol. 4.* New York: Academic Press, 103-194.

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

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial ordering mechanism in speech production. In W. E. Cooper & E. C. T. Walker (eds.), *Sentence Processing: Psycholinguistic Studies presented to Merrill Garrett.* Hillsdale, NJ: Erlbaum, 295-342.

Studdert-Kennedy, M.G. (1998). The particulate origins of language generativity: From syllable to gesture. In Hurford, J. R., Chris Knight, C. & Studdert-Kennedy, M.G. (eds.), *Approaches to the evolution of language.* Cambridge: Cambridge University Press, 202-221.

Studdert-Kennedy, M. G. (2000). Evolutionary implications of the particulate principle: Imitation and the dissociation of phonetic form from semantic function. In Knight, C, Studdert-Kennedy, M.G. & Hurford, J.R. (eds.), *The evolutionary emergence of language: Social function and the emergence of linguistic form,* 161-176. Cambridge: Cambridge University Press.

Van Hooff, J.A.R.A.M. (1967). Facial displays of the catarrhine monkeys and apes, in Morris, D. (ed.), *Primate Ethology.* London: Weidenfield and Nicholson, 7-68).