Davis, B.L. & MacNeilage, P.F. (2004). *Primatologie,* Volume 6, pp. 305-328.

The frame/content theory of speech evolution: from lip smacks to syllables

Barbara L. Davis and Peter F. MacNeilage

The University of Texas at Austin

The aim of this paper is to illustrate the scope and claims of the Frame/Content theory (MacNeilage & Davis, 1990, 2000; MacNeilage, 1998). We will explore the implications of Frame/Content for consideration of ontogeny and phylogeny as well as modern language structure. At a fundamental level, we propose a biologically based view of the vocal-auditory medium founded on biomechanical capacities and rooted in deep evolutionary time.

Basic to our conceptualization is 'embodiment', the philosophical perspective whereby "Mental activity and underlying brain activity cannot be understood outside the context of bodily activities" (Clark, 1997, pg 23). This construct has been widely explored in disciplines as diverse as philosophy (Johnson, 1997), linguistics (Lakoff & Johnson, 1999), cognitive science (Varela, Thompson & Rosch, 1997), dynamic systems theory (Thelen & Smith, 1994), neuroscience (Edelman, 1992), and artificial intelligence (Clark, 1997). For speech production, a fundamental corollary of embodiment is the ability of speakers to produce rhythmic strings of syllables – speech's solution to the general problem of serial order in behavior (Lashley, 1951).

In our view, earliest speech emerged as an adaptation of the available movement capacity of the vocal apparatus in hominid speakers. Biomechanically simple serial vocal patterns paired with meaning in early protoword forms were thus made available for cultural transmission in service of interpersonal and community-wide communication. We propose that this modern speech form was originally founded on cultural invention, emerging from early speakers' socially driven needs interact with one another and, over time, to develop a larger message set for communicating increasingly complex messages. The syllable, a basic form employed by modern speakers is ubiquitous in contemporary language (Maddieson, 1984). Contemporary speech forms reflect the endpoint of the evolutionary process. With this endpoint in mind, consideration of the auditory vocal medium and the mechanisms supporting employment of that medium for linguistic communication should be central to development of a phylogenetic perspective.

The Frame/Content theory related to the auditory-vocal medium proposes that consonants and vowels ("content" elements such as "r" or "w") are placed into syllable structure "frames" which obey the ordering regularities of each of the 5,000 or so modern languages (MacNeilage & Davis, 1990, 2000; MacNeilage, 1998). These ordering regularities are diverse, ranging from the simple CV forms in Hawaiian to the complex forms of English (i.e. up three to four consonants can occur at the beginning and at the end of a syllable form). A basic duality is proposed between syllable related structures ("frames") and segment related structures ("content"). The "frame" in modern adult speakers is based on a pre-motor syllabic receptacle for consonant and vowel placements. Pre-motor refers to the planning and organization process preceding actual production of speech by the peripheral mechanisms such as the tongue, lips, and jaw (Levelt, 1999).

This duality proposal is not unique to the frame/content theory and has been widely accepted based on the psycholinguistic study of speech errors in adults (Levelt, 1991). Studies of speech errors have shown that when individual consonants and vowels are misplaced in an otherwise correct utterance, they go into the same positions in syllable structure where they originated. Syllable-initial consonants reverse with syllable-initial consonants, as in “gad birl” for “bad girl”. Vowels reverse with vowels: as in. “bag cit” for “big cat”. Medial consonants reverse with medial consonants, as in “bimest cagel” for “biggest camel”. Independent control of the individual segmental or “content” elements is instantiated in the pre-motor phase of speech planning previous to carrying out movements of the oral articulators that can operate independently of the jaw in mature speech.

In infants, canonical babbling represents the first onset of speech-like behavior, in that it contains consonant and vowel-like alternations (i.e. “baba” or “dadada”). It also is rhythmic in “sounding like speech” to an adult listener, even though the infant does not yet have words (see Oller, 2000 for a review of canonical babbling). In contrast to adult speakers, where the frame is an aspect of *pre-motor* planning, the frame in infants at the onset of speech-like vocalizations in canonical babbling is based in actual speech movements. The infant frame is an emergent product of rhythmic close and open jaw cycles, a biomechanically based operation of the peripheral speech mechanism. These pendulum-like jaw cycles are repeated, resulting in a chopping into speech-like sound qualities emerging from the rhythmic close and open alternations of the infant’s jaw accompanied by phonation. Phonation is based on the fundamental frequency of vibration of the vocal folds. A consonant percept (e. g. “b”) results from the close phase of the jaw cycle and a vowel percept (e. g. “a”) from resonance properties is enabled by the open phase. No autonomy of speech

articulators such as the tongue, lips, and soft palate (the structure used to close off the nasal cavity, necessary for oral speech sounds such as “b”) is required for producing an adult listener perception of speech-like rhythmic vocalization. Action in the time domain is a basic to this conceptualization. Thus earliest vocalizations emerge from peripheral motor capacities for achieving rhythmic close and open jaw movements, the motor frame. We have argued elsewhere that this earliest step is reflective of self-organization in a complex system (MacNeilage, Davis, Kinney & Matyear 2000a).

How is this conceptualization of change for the modern syllable from rhythmic movements of the jaw in infant vocalizations early in development to a pre-motor planning status in adults relevant to consideration of the phylogeny of the auditory-vocal medium? We would contend that the *pre-motor* frame constraining speech errors in modern speakers evolved from the *motor* frame of mandibular oscillation in early users of a vocal communication mode. Most generally the “frame” had its origin in ingestive mandibular cyclicities available to early humans and was subsequently adapted for communication by pairing with phonation. Mutual exclusiveness of consonants and vowels likely emerged from the mutual exclusiveness of close and open cycles in early rhythmic alternations based on elevation and depression of the mandible. In this sense, the first speech-like sound patterns of infants in babbling and first words may provide a window on earlier stages of the evolutionary process that terminate in adult speech patterns in contemporary languages. Glimpses of the origins of the pre-motor frame may be found in the earliest rhythmic jaw oscillation patterns in infant speech-like output.

The infant develops independent control of the coordinated articulators within in the time constraints necessary for maintaining rhythmic behavior. This process of gradual

mastery of independent articulator movements within utterances enables accurate productions of specific consonant and vowel elements (such as “s” or “m” or “i”) in the infant’s ambient language by 4-5 years. In our perspective, these “content” elements (consonants or vowels) are gradually differentiated from the jaw-based “frame” to become separate entities as the infant acquires increasing control over the coordination of articulators in vocal sequences. In particular, this development is primarily the result of introducing additional on-line variation (modulation) in the basic frame cycle, as well as on-line variation in the positioning of the tongue, soft palate, and lips.

As the tongue becomes relatively more independent of the jaw, the infant initiates place variegation (i.e. moving the tongue within consonant-vowel sequences) thus reducing syllable reduplication (“baba” becomes “bado” for “bottle”). Control over soft palate closure is evidenced by growth in alternation of nasal consonants (i.e. “m”, where the soft palate is open) and orals (i.e. “b”, where the soft palate is closed) in sequences rather than pervasive nasal (“mama”) or oral (e.g. “baba”) quality throughout the utterance. Capacity for production of “mama” or “baba” with an open or closed soft palate throughout develops into the ability to change the position of articulators (e. g. change in soft palate closure to produce “bam” or “don”) within the utterance.

In addition, during this later period, maturing perceptual and cognitive capacities enable infants’ increasing recognition and learning of salient sounds and sequencing regularities in the ambient language input (see Jusczyk, 1997, for a review of early perceptual maturation in human infants). The perceptual-cognitive component is thus essential to learning of the precise nature of “content” elements specific to the infant’s own ambient language across the course of acquisition.

At a fundamental level, speech is *output*. As a result, perceptual and cognitive capacities must be viewed as "in service" of the production system abilities of speakers (Davis & MacNeilage, 2000). In later stages of ontogeny, the production system matures to enable simulation of more diverse speech models from the ambient language in tandem with maturing perceptual and cognitive capacities. Understanding of the course of phylogeny, we have argued, is well served by recognition of the nature of self-organizing forces shaping early motor-based "frames". Recognition of the early importance of the jaw cycle gives a powerful starting point for understanding the emergence in phylogeny of initial motor-based vocal forms. The gradual emergence of "content" as motor, perceptual, and cognitive capacities mature provides a potential window on the step-wise process for understanding gradual emergence of complexity in phylogeny as well, motivated by external social and cultural pressures for complex message transmission.

In summary, first speech-like sound patterns of infants in babbling and first words may provide a window on earlier stages of the evolutionary process terminating in speech patterns observable in modern languages. Glimpses of the origins of the modern pre-motor frame in contemporary speakers may be found in the earliest mandibular oscillation patterns characteristic of infant speech output. Observable linguistic patterns spring from the context in which real languages are learned, and, more important, in which they evolved. Like any other set of conventions, linguistic conventions are shaped by the social and cultural situations in which they originated. This possibility, based in the potential function of simple forms for communication in early hominid speakers, allows us to suggest an answer to the question of *why* consonants and vowels never occupy each other's positions in syllable structures in modern speech errors (e.g. "abstract" never becomes "bastract"). The

explanation for this conceptually important autonomy can be found in the requirement for mutually exclusive movements for consonants and vowels from babbling onward involving elevation of the mandible for consonants versus depression of the mandible for vowels. Phylogenetically, the basic consonant and vowel movements were antagonistic. They were autonomous, resulting in their separate status in the evolution of an eventual pre-motor control program. Thus from an embodiment perspective, a body-to-mind progression was responsible for the present day existence of cognitive-motor CV frames.

Pre-speech Origins

What might have been the phylogenetic origin of these rhythmic mandibular cyclicities prior to their implementation in early hominid users of an oral communication medium? MacNeilage (1998) has suggested that the basic open-close mandibular "frame" for speech had its origin in ingestive mandibular cyclicities (e.g. chewing, sucking, licking). The mandibular cycle is an oral movement pattern basic to all mammals. From the frame/content perspective the consonants and vowels produced by modern speakers likely emerged from early rhythmic alternations based on elevation and depression movements of the mandible available to paleomammals by about 200 million years ago. One piece of evidence for support this claim is that an aspect of the cortical component of the extrinsic production system, partly involving a homolog of Broca's area, is the main cortical control center for mammalian chewing **(Woolery, 19??)**. An intermediate form in the progression from ingestive cyclicities to CV syllables and intersyllabic sequences can be proposed based on visuofacial communicative forms such as lipsmacks, tonguesmacks and teeth chatters common in other modern primates before it was used for speech (Redican, 1975; Van Hooff,

1962; 1967)). In these intermediate forms, the rhythmic cycle was exapted for visuofacial communication. Each of these forms involves mandibular oscillation, and most likely shares a single rhythm generator at the neural level. Lipsmacks are an important communicative accompaniment to manual grooming (Redican, 1975). As such they may have evolved a communicative function as anticipatory actions for putting grooming particles into the mouth (Van Hooff, 1962; 1967).

The earliest protosyllabic vocalizations in hominids may have co-opted two output capacities already co-ordinated in mammals; phonation, a laryngeal function based on vocal fold oscillation, and articulatorily based syllable-like alternations between open and close mouth configurations. As we have emphasized, these characteristics are evident in earliest phases of canonical babbling (Oller, 2000), in vocal sequences with syllable-like timing regularities based on jaw open and close cycles perceived as "speech-like" by adult listeners. Infants are recapitulating phylogeny at the first appearance of speech-like behavior when they co-ordinate rhythmic mandibular oscillations with phonation at the onset of canonical babbling, several months before communicative words begin to appear.

Neural Organization

Crucial to understanding control mechanisms for the frame/content theory is consideration of underlying neural organization. In this regard we assert that the behavioral duality of frames and content is paralleled by a dichotomy in brain organization. Relative to the *frame* component, data on several classes of neurologically impaired adults is supportive of our assertions. These patients involuntarily produce strings of reduplicated consonant-vowel syllables – the most basic syllable type (e.g. "babababa"). Their pathology is either a

result of artificial electrical or seizure-related interference to an area of medial premotor cortex (Supplementary Motor Area, SMA) or a form of global aphasia releasing the SMA from intentional control. The SMA is the main cortical component of a generalized "Intrinsic" (self-generated) movement control subsystem (Goldberg, l985). Thus the frame can be seen as fundamental to intrinsically generated movement sequences in speech that is conserved in damaged neural tissue. This conservation in the face of extensive neural tissue damage may be fruitfully considered as conservation of fundamental movement capacities available to earliest hominid speakers.

It should be noted parenthetically that in recent years a distinction has been made in medial premotor cortex between a rostral area called the pre-SMA and a more caudal SMA proper, still termed SMA. Kennerly, Sakai and Rushworth (2004) suggest that "for overlearned sequences … the pre-SMA is primarily concerned with the initiation of a sequence or sequence-chunk" (p. 978) In contrast, they suggest that the SMA "may code for intervals between specific movements within a sequence, and the rank order of sequence movements." (p. 979) Consistent with the present emphasis on the Intrinsic role of medial premotor cortex, these authors note that the "medial premotor areas, including the pre-SMA, seem less essential when a specific movement is to be selected on the basis of an association with a sensory cue." (p. 989)

In contrast, the main cortical component of the subsystem producing *content* is found in the frontal lobe regions of ventral lateral, opercular and insular frontal premotor cortex, including parts of Broca's area. These cortical regions participate in an "Extrinsic" (externally driven) movement control subsystem (Goldberg, 1985). This extrinsic component mediates an evolving mimetic capacity - partly involving the mirror neurons discovered in ventral lateral premotor

cortex by Rizzolatti and colleagues (Rizzolatti, Fogassi and Gallese, 2001). This class of "mirror neurons" has been found in the monkey homolog of Broca's area. Mirror neurons discharge when a monkey performs an action such as grasping. More importantly, mirror neurons also discharge when a monkey observes others performing the same action. Rizzolatti and colleagues have now shown that there are mirror neurons in this region involved in both ingestive behaviors and visuofacial communicative behaviors including lipsmacks (Ferrari et al., 2003). Thus for both of the neurons most crucial for our thesis that lipsmacks may have been precursors to proto-syllables, input and output functions were both communicative.

Cultural Context

Another perspective on origins providing evidence for the frame/content theory can be found in Merlin Donald's (1991) proposal of "mimesis" or gestalt social simulation. Gestault social simulation is demonstrated in modern cultural rituals such as dance or sport where participants interact relative to movements of the other participant's body movement and rhythms. Donald has proposed that mimesis formed a crucial step toward gaining cognitive instantiation of language. Original mimetic capacities observed in early cultural-social rituals are seen as initially dependent on overall motor capacities, a generalized adaptation of the available movement repertoire of the early hominid. Mimetic sequences reflected a generalized ability to refine actions based on cognitive-motor linked representational capacities for simulating external action with one's own body. Crucially for the basic tenets proposed in the frame/content theory, Donald proposes that the ability to make the sound patterns for earliest word forms in a social-communicative context was likely a necessary precondition for making these words, rather than evolving after the cognitive

instantiation of word meanings based on conceptual-linguistic capacities (e. g. Bickerton, 1990). The addition of phonation to these mandibular cyclicities, yielding protosyllabic sequences, could have been a facilitator of the transition to vocal grooming as suggested by Dunbar (2001). Dunbar proposes a socially motivated scenario propelling the early hominid from social grooming to vocal calls related to the costliness of maintaining social alliances via social grooming.

Speech Forms

Consideration of modern language forms and the relationships between infant babbling/early word patterns and those found in modern languages afford another set of datum for considering the origins of communicative vocalizations in early hominids. The possibility that patterns in babbling and early words are important to understanding of language origins is increased by the persistence of these preferences in modern languages. The CV syllable, the favored syllable type in infant vocalizations is the only universal syllable type in languages (Bell & Hooper, 1984). Consonants favored in babbling (i. e. oral and nasal stop consonants) are highly frequent in languages (Maddieson, 1984). These sound qualities predominate in the phonemic repertoire of languages with small systems (<15 phonemes) containing a few segments characterized as articulatorily "simple" (Lindblom & Maddieson, 1988; Lindblom, Krull & Stark, 1993). In addition to providing a rationale for these general parameters of modern linguistic form, the frame/content theory produces precise predictions for the intrasyllabic and intersyllabic serial structure of vocalizations in infants as well as in modern languages. These predictions can be productive for considering the structure of earliest vocalization forms.

Intrasyllabic Patterns: The frame/content theory produces an intrasyllabic hypothesis predicting strong associations between close and open phases in mandibular oscillation cycles resulting in three regularities of labial, lip, consonants with central vowels (e. g. [ba]), coronal, tongue tip, consonants and front vowels (e. g. [di]), and dorsal, tongue back, consonants with back vowels (e. g. [ku]). From a biomechanical perspective, these CV co-occurrences within syllables seem straightforward. The two lingual patterns (coronal-front and dorsal-back) may result from a basic physical property of matter – inertia. The tongue has a tendency to remain in a similar position in the mouth within a syllable rather than moving from one position to another. In addition, like the tongue, the soft palate tends to not change position during early utterances making the entire babbling episode nasalized if the soft palate is at rest in its open position [mãmãmã] and non-nasal if the soft palate is elevated throughout the episode [bababa] (Matyear, MacNeilage & Davis, l998). Thus, the patterns observed for labial and coronal consonants tend to be similar whether the consonants involved are oral ([b], [d]) or nasal [m], [n]), except for some nasal resonance effects on vowels in nasal contexts.

A variety of studies have tested these intrasyllabic predictions of the frame/content theory made in an original case study of early speech production made in 1990 (Davis & MacNeilage, 1990). In addition, Davis & MacNeilage (1995) studied 6 English-learning infants during babbling. The three predicted intra-cyclic CV associations, involving stops, nasals, and glides, were confirmed at significant levels in all infants (18 total predictions for 6 infants). Only 9 out of 36 above chance associations were found in CV associations that were not predicted. Gildersleeve-Neumann, Davis and MacNeilage (2000) also considered lower frequency consonants (fricatives, liquids, and affricates, often very late to appear in

typical speech development) in four of the infants previously studied in babbling. With minor exceptions, CV co-occurrence patterns similar to the patterns for stops, nasals, and glides were found (i.e. coronal consonants with front vowels and labial consonants with central vowels). A subsequent study of 10 infants during the single word period, 4 of whom were also studied during babbling (Davis, MacNeilage & Matyear, 2002) revealed the three predicted CV co-occurrence trends in 25 out of 28 instances, while non-predicted trends only occurred in 17 out of 58 cases (Chi Square, p<. 01). These results indicate that the intrasyllabic co-occurrence trends found in babbling do not appear to resolve during the single word lexical acquisition period in typically developing infants in an English environment.

A second source of behavioral evidence for the generality of these intrasyllabic predictions can be found in studies considering these trends in infants in language environments other than English (Davis & MacNeilage, 2002). Table 1 displays results of a survey of available studies. Based on X^2 analysis, results are expressed as a ratio of observed-to-expected frequencies for each cell (Expected value is 1.0). Expected frequencies were derived from the overall frequency of the particular consonant and vowel types in the corpus. Thus, if 80% of all consonants were coronal consonants, and 40% of the vowels were front vowels, then the expected frequency of coronal consonants with front vowels would be .8 X .4 = .32 or 32% of all CV pairings. Studies of Japanese and Swedish (five infants in each language) from the Stanford Child Phonology Project (Vihman, 1992) are included. Seven Ecuadorian-Quichua infants (Gildersleeve-Neumann, 2001) were studied in babbling and in first words. Two Brazilian-Portuguese-learning children in the single word period (Teixeira & Davis, 2002) were analyzed. One Italian infant (Zmarich, & Lanni, l999) was followed in

babbling and in the single word period. Serbian-English bilingual twins were studied during babbling (Zlatic, MacNeilage, Matyear & Davis, 1997). Five Korean-learning infants (Lee, 2000) were also studied across the babbling and single word periods.

Table 1.

Mean Observed-to-Expected Ratios For The Three Predicted Consonant-Vowel Co-occurrence Patterns In Infants: Swedish, Japanese, Ecuadorian-Quichua, Brazilian-Portuguese, Italian, Serbian, And Korean Environments.

	Swedish			**Japanese**			**Quichua**			**Braz-Port.**			**Italian**			**Serbian**			**Korean**		
N	5			5			7			2			1			2			5		
	C	**L**	**D**	**C**	**L**	**D**	**C**	**L**	**D**	**C**	**L**	**D**	**C**	**L**	**D**	**C**	**L**	**D**	**C**	**L**	**D**
F	1.2			1.2			2.6			1.0			1.3			1.2			1.7		
C		1.2			1.2			1.6			1.2			1.4			1.3			1.3	
B			1.4			1.1			2.1			1.8			1.0			NA			2.0

C, L, D = **C**oronal, **L**abial, **D**orsal Place of Consonant Articulation

F,C,B = **F**ront, Central, **B**ack Dimension of Vowel Tongue Placement

In 27 infants across 7 languages, CV associations were broadly confirmed with few exceptions and no consistent counter trends. Confirmation of the predicted associations in the majority of these infants in diverse language environments supports the predictions of early dominance of a jaw oscillation based “frame” as a general characterization of intrasyllabic properties in babbling and early speech vocalizations.

Studies of hearing-impaired populations show that normal auditory perceptual input is an important causal factor in the production of frames. Infants with profound or even severe hearing loss sometimes do not babble (e.g. Oller & Eilers, 1988; Oller, Eilers, Bull & Carney, l985). In these infants, there is a marked skewing in the distribution of sounds produced toward labial and nasal consonant qualities and neutral vowel positioning. This asymmetry has been attributed to a differing balance between visual and auditory information, heavily favoring the visual modality in the relative absence of auditory input.

To explore our expectations of the role of the production-perception interface in motivating the CV co-occurrence constraints, we studied one profoundly hearing impaired infant who received a cochlear implantation procedure at 25 months (McCaffrey, Davis, MacNeilage & von Hapsburg, 2000). Pre-implant data was derived from two sessions 3 and 1 months before implant, and from sessions 2, 7 and 9 months post-implant. Pre-implant, the infant was typical of profoundly hearing impaired children in producing relatively few syllabic utterances, mostly labial-nasal stop consonants and neutralized vowels. There were too few coronals and front vowels and dorsals and back vowels to allow a comprehensive statistical analysis of CV co-occurrence patterns. Post implant, both labial-central and coronal-front CV co-occurrences were plentiful. Combining 7 and 9-month post-implant data, the observed-to-expected ratio of labial-central CVs was 1.14 and the ratio for coronal-front CV's was 1.62. The coronal-front value was much higher than those observed in hearing infants. Both effects were significant. These results on a young CI infant confirm the robust nature of the labial-central and coronal-front CV co-occurrence patterns. When the infant had enough canonical syllable type vocalizations post-implant for analysis, she

produced the predicted intrasyllabic preferences we have identified based on characteristics of the production system found in hearing infants.

In considering the relative explanatory power of patterns founded on the nature of the production system in the face of diminished auditory perceptual input, we found that when enough syllables were produced to analyze, predicted CV co-occurrences (indicating a relative lack of articulator movement within syllables independent of the jaw) predominate. This finding underscores the strength of the frame/content theory predictions, even in cases of differing perceptual input sufficient to strongly compromise typical developmental sequences.

Languages

The strength of these patterns in ontogeny moved us to consider the possibility of persistence these patterns in modern language forms. Two sets of data each containing a different set of five diverse languages, originally described by Janson (1968) and Maddieson & Precoda (1990) were re-analyzed (MacNeilage & Davis, 1993). We found evidence of common tongue tongue-fronting and tongue-backing but not of the co-occurrence constraint involving labial consonants and central vowels. In a further study including all English consonants (Davis, MacNeilage & Matyear, 1999) based on analysis of 33,654 lexical types in the *Shorter Oxford English Dictionary* (*SOED*; Quinlan, 1992) we found the coronal-front and dorsal-back, but not the labial-central associations present at significant levels. MacNeilage, Davis, Kinney & Matyear, (2000b) also found all three intrasyllabic properties to be largely present in a dictionary analysis of a group of 10 languages (i. e. English, Estonian, French, German, Hebrew, Japanese, New Zealand Maori, Quichua, Spanish and Swahili), suggesting that they are basic to the operation of the speech production system in

general and are retained in modern languages, rather than simply being transient aspects of early speech development. Rousset (2003) investigating data on 14 languages available on the ULSID (UCLA Lexical and Syllabic Inventory Database; Maddieson, 1992) database, created incorporate genetic and geographical diversity into language typology study. Their analysis confirmed the three CV co-occurrence predictions in these languages as well.

These results raise the possibility that the tongue-based constraints postulated for babbling and early speech may be extremely fundamental to modern speech movement systems. They indicate the importance of a physical principal of movement inertia, in contrast to a perceptual distance hypothesis (e. g. Kawasaki, 1982) in the formation of CV patterns in languages. The coronal-front and dorsal-back CV co-occurrence patterns involve the same mandibular oscillation pattern as the labial-central frames in infants. A non-resting (fronted or backed) tongue position is characteristic in these two patterns. Consonants and adjacent vowels tend to share a similar tongue position in the front-back axis in coronal and dorsal CV forms. In labial consonantal contexts, lack of lingual engagement for the consonant leaves the vowel potentially freer to vary during the vocalic portion of the syllable to enhance diversification. Some languages have taken advantage of this potential and incorporated it into the structure of modern word forms. Thus, the labial-central association is somewhat weaker than the lingual associations.

In addition, another structural regularity is revealed in analysis of infant acquisition data in babbling and first words. In infants, the same biomechanical constraint against tongue movement within a CV sequence is equally evident in VC sequences. This regularity would be expected from infants' tendency to simply repeat the same frame (Davis et. al., 2002). However, both our studies of 10 languages (MacNeilage et al, 2000) and the study of 14

languages by Rousset (2003) indicate that the basic biomechanical constraint indicated by these co-occurrence patterns is overcome in languages. We have found none of the three patterns of co-occurrence in VC sequences in language analyses. In languages, the boundary between syllables tends to occur between a V and a following C. Absence of the VC co-occurrence patterns accompanies the regularity in found languages whereby successive syllables strongly tend to differ from one another. This difference suggests that reversal of the infant (and earlier hominid) trend involves overcoming a natural biomechanical tendency. This basic tendency is overcome in the service of developing the syllable as an independent structural entity in speakers. This change is at the vortex of biological and social-cultural forces in phylogeny supporting diversification of vocal structure to achieve message diversification. Our supposition, based on this difference in vocal form between modern infants and adults is that earliest vocal forms would likely have shown the VC as well as the CV co-occurrences. That is, early speakers began by implementing a vocal communication mode we can observe in modern ontogeny.

Intersyllabic Patterns: The frame/content theory also produces predictions about intersyllabic regularities in infant babbling and early words. As we have noted, consonants and vowels in successive syllables tend to be the same (e.g. [baba]). This pattern is termed ‘reduplication’ (Oller, 1980). Reduplication can be viewed as based on the same simple biomechanical foundation. An additional biphasic cycle of oscillation of the mandible is produced virtually identical to the initial cycle. No changes in other articulators are necessary in order to produce this syllabic “reduplication”.

What about languages? Languages show the opposite pattern to the reduplication patterns favored in infants. In our survey of dictionary counts in 10 languages; MacNeilage

et. al., 2000b), language patterns uniformly showed below chance levels of consonant reduplication. We have postulated that VC co-occurrence constraints emerged in service of the social-cultural pressure to diversify messages (MacNeilage et. al., 2000b). Introduction of variegation in place of reduplication in modern languages represents an aspect of language structure where languages reflect the pressure to overcome movement inertia in serial vocalization patterning in support of diversification for message building complexity.

However, infants also produce some variegation in vocal sequences from the onset of babbling (e. g. Mitchell & Kent, 1991). The frame/content-based prediction is that most of the change in infant variegated sequences might also be attributable to the mandible via variation across syllables in the amplitude of close-open dimensions. From this perspective, most intercyclical vowel variegation would be predicted in vowel height over front-back change (e. g. [daedi] over [daedu]) related to lack of tongue movement in the horizontal (back and forward) dimension within sequences. Consonant manner variegation related to degree of closure and tongue elevation would predominate over *place* variegation involving front-back tongue movement across syllables (e. g. [bawa] over [baka]).

We have confirmed the prediction that most intersyllabic variegation is attributable to amplitude of mandibular excursions in babbling and the single word periods. This prediction was confirmed in utterances involving stop, nasal and glide consonants during babbling in 6 English learning infants (Davis & MacNeilage, l995). We have also found that the tendency to favor manner and height variegation persists across the single word period in as study of 10 English-learning infants (Davis et al., 2002). As with the within syllable CV co-occurrences, babbling episodes involving later developing consonants such as fricatives, affricates and liquids (Gildersleeve-Neumann, et al., 2000) also demonstrate these intersyllabic patterns.

Our prediction has also been tested and confirmed in the single word period in Equadorian-Quichua (Gildersleeve-Neumann, 2000), Korean (Lee, 2003), and Brazilian-Portuguese (Teixeira & Davis, 2002). Thus, we would assert that most of the variance in early variegated sequences is being produced by variation in amplitude of the mandibular cycle across successive close and open phases in vocalization stings that are variegated. In addition, it is important to note that there is no obvious tendency to repeat the same variegation pattern across utterances in babbling or in early words. This lack of consistency in the nature of variegation within a sequence in an utterance suggests that it might not be under the intentional control of infants in this period.

This infant patterning introduces a complication. Modern adult speakers and languages reduplicate at below chance levels within utterances. The question arises of how infants begin to move into the intersyllabic complexity noted in mature speakers and modern languages. Ingram (1974) noted that consonants in infants producing their earliest words tend to begin the word with sounds more forward in the mouth (i.e. lip sounds like “b”, or tongue tip sounds like “d”). Subsequent consonants are produced further back than this initial consonant closure. Since infants produce few dorsals (i.e. “k” or “g” as in “kick” or “go”) in early words, the main manifestation of this pattern of beginning closure at the front of the mouth in a word is a labial-vowel-coronal (LC) sequence (e. g. [bid] for “bead”).

We examined 7 data reports on 5 diverse language environments. Analysis of CVCV and VC word patterns revealed that the LC tendency predominated in 21 out of 22 infants studied (MacNeilage & Davis, 2000). Two infants sometimes changed the CL order in the target word to the preferred LC sequence (as in “top” - > “pot”) (Jaeger, l997; Macken, l978). In another study of 10 English-learning infants producing single words, nine of the ten showed the LC pattern and

the other showed no trend (MacNeilage, et al, 2000a). The mean LC: CL ratio for sequences was 2.55.

From our perspective, labial consonants and central vowels (e. g. “ba”) may be easier to produce than coronal consonant initiations that require tongue-front presetting (e.g. “di”). At the first word stage when the infant has to interface a vocal episode with a specific lexical concept rather than simply producing a vocalization string we see a trend towards increasing use of labial-central initiations, followed by the coronal tongue front movement to achieve place variegation. The initiation of action is a complex function in motor systems (MacNeilage & Davis, 2001). The more complex requirements for output in connecting vocalizations to meaning in early words make the functional load on the motor system greater. In this context, responding to the increased functional load, the infant appears to be initiating with an easier lip or labial closure and adding subsequent tongue movement or coronal closure, thus increasing complexity, after the utterance is generated.

As with the intrasyllabic CV co-occurrences and reduplication patterns, a crucial piece of data is the nature of place variegation in adult speakers and languages. Our dictionary analysis of 10 languages (MacNeilage et al, 2000b) indicated that the LC preference was present in 9/10 and was significant in 8/10. The mean ratio of LC to CL sequences was 2.23. We also examined labial-dorsal (LD) and dorsal-labial (DL) sequences in the corpora. LC was twice as frequent as the other five possibilities (CL, DL. LD, CD, and DC). Mean frequencies of these five other potential sequences ranged from 13.1 to 16.2%. Languages show the infant excess of LC over CL strongly, indicating another site of evolutionary conservation of basic propensities based on biomechanics of movement systems, reflective of self-organization of multiple agents in a system to produce complex output.

Does the progression we have described provide insight into the nature of early vocal sequences first hominid speakers? In phylogeny, perceptual support from already existing adult word targets was not present. Thus initial pressure was towards creating rather than simulating lexical diversity, as no external models were available for early speakers. However, as with infants, we assert that this early place change pattern is reflective of self-organization in a complex biomechanical system thus not reflective of early learning.

Summary

Fundamental to application of the frame/content theory in consideration of phylogenetic origins of modern speech forms is our proposal of biomechanical basis for the operation of the speech production system. Our proposal is based on the suggestion that prototypical speech-like vocal communication based on the jaw cycle and accompanying phonation was co-opted from earlier functional routines of chewing, sucking, and licking which are rhythmic and involve open and close jaw cycles. These rhythmic "frame" behaviors are instantiated in neural tissue in the SMA, the main cortical component of a generalized "Intrinsic" (self-generated) movement control subsystem. The frontal lobe regions of ventral lateral, opercular and insular frontal premotor cortex, including parts of Broca's area, the main cortical component of the subsystem underlie "content" based on an "Extrinsic" (externally driven) movement control subsystem.

Speech ontogeny entails an initial frame stage involving basic capabilities available to the production system. Earliest frames provide evidence of self-organizing processes in a complex system rather than learning. Our postulation of a parallel frame stage for the phylogenetic origins of speech is suggested by evidence that some of the main features of the intrasyllabic frame stage in infants are also present in syllables in modern languages As we have noted, characteristic

intra- and inter-syllabic frame patterns found in infants are not clearly learned from the environment, as they are typically stronger that the associations found in languages and are present in some instances when they are not found in the ambient language (i.e. in the case of labial-central associations, not consistently present in languages). Maintenance of these characteristics of early ontogeny in mature language users (i.e. the CV co-occurrence patterns, the LC place change effect) indicates that they may be extremely basic to the operation of the movement control system.

Later stages of content emergence reflect learning of ambient language regularities in modern infants and social-cognitive pressures to increase the message set in early hominid speakers. This diversification pressure would have necessarily been based on cognitive prerequisites including knowledge and memory. Our hypothesis is generally consistent with Donald' mimesis proposal (1998) that movement capacities were available first for evolutionary tinkering into an eventual representational system linking sound-making capacities to meaning for social communication. Evolution may likely have preserved functional aspects of the internal structure of the jaw cycle (i.e. intrasyllabic regularities) while elaborating over time with additional adaptations (place variegation) seen in increase of independent articulator movements within syllables in the service of increase in message complexity.

We have pointed out a number of similarities and differences between infant babbling and early speech and sound patterns of languages to provide performance-based explanation for observed patterns. A key principle of the frame/content theory is that sound patterns of words are a result of two sometimes conflicting forces – production constraints, and the need for perceptual distinctiveness. With regard to the evolution of *pre-motor* language knowledge evidenced in modern speech errors in adult speakers, our conclusion is that these capacities were

not given in advance by a fortuitous mutation, providing a pre-motor structure that just happened to instantly resolve the conflict between the productive and perceptual pressures. Instead a body-to-mind progression implemented in extrinsic and intrinsic control systems arising from precursors observed widely in mammals provides a rich matrix for evolving speech from generally available non-speech capacities in early hominids.

References

Bell, A. & Hooper, J.B. (1978). Syllables and segments. North-Holland, Amsterdam.

Bickerton, D. (1990). Language and Species. University of Chicago Press: Chicago.

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

Clark, A. (1997) Being there: putting brain, body and world together again. MIT Press, Cambridge .

Davis, B.L. & MacNeilage, P.F. (1990). The acquisition of vowels: a case study. Journal of Speech and Hearing Research ,33,16-27.

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.; Matyear, C. (1999). Intrasyllabic patterns in babbling and early speech. Proc. 14th Int. Cong. Phonet. Sci., San Francisco.

Davis, B.L. & MacNeilage, P.F. (2000). An embodiment perspective on the acquisition of speech perception. Special Issue. Phonetica. 57, 229-241.

Davis, B.L. & MacNeilage, P.F. (2002). The Internal Structure of the Syllable: An Ontogenetic Perspective on Origins in Givon, & Malle, (Eds.) The Rise of Language out of Pre-language. Benjamin

Davis, B. L., MacNeilage P.F. & Matyear, C. (2002). Acquisition of serial complexity in speech production: A Comparison of Phonetic and Phonological Approaches to First Word Production. Phonetica, 59, 75-107.

Donald, M. (1991). Origins of the Modern Mind. Harvard University Press, Cambridge, MA.

Dunbar, R.J.M. (2001). Brains on two legs: Group size and the evolution of intelligence. In deWaal (ed.) Trees of Origin, Harvard University Press: Cambridge MA.

Edelman, G.W. (1992). Bright Air, Brilliant Fire: On the Matter of Mind. Basic Books, New York: New York.

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

Gildersleeve-Neumann, C.E., Davis, B.L. & MacNeilage, P.F. (2000). Contingencies governing production of for fricatives, affricates and liquids in babbling. Applied Psycholinguistics. 21, 341-363.

Gildersleeve-Neumann, C.E. (2001). Constraints on infant speech acquisition: a cross-language perspective. Ph.D. dissertation, The University of Texas at Austin, Austin, TX.

Goldberg G. (1985) Supplementary motor area structure and function: Review and hypothesis. Behav Brain Sci, 1995, 8,: 567-616.

Janson T. (1986), "Cross-linguistic trends in the frequency of CV sequences. Phonology Yearbook, 3, 179-195.

Johnson, M. H. (1997). Developmental Cognitive Neuroscience. Blackwell: Cambridge.

Jusczyk, P. W. (1997). The Discovery of Spoken Language. MIT Press: Cambridge, MA.

Kennerly, S.W., Sakai, K. and Rushworth, M.F.S. (2004) The organization of action sequences and the role of the pre-SMA. Journal of Neurophysiology, 91, 978-93.

Kawasaki, H. (1982). An acoustical Basis for the Universal Constraints on Sound Sequences. Ph.D. dissertation, The University of California, Berkeley.

Lakoff, G. & Johnson, M. (1999). Philosophy in the flesh: the embodied mind and its challenge to western thought. Basic Books, New York.

Lashley, K. (1951). The problem of serial order in behavior; in Jeffress; Cerebral mechanisms in behavior. Wiley: New York, 51-66.

Lindblom, B.; Krull, D.; Stark, J. (1993). Phonetic Systems and Phonological Development in Boysson-Bardies, de Schoen, Jusczyk, MacNeilage, Morton; Developmental neurocognition: speech and face processing in the first year of life. Kluwer: Dordrecht, 399-409.

Lindblom B. (2000). The Interplay between Phonetic Emergents and the Evolutionary Adaptations of Sound Patterns, Phonetica, 57, 297-314.

Lindblom, B. & Maddieson, I. (1988). Phonetic universals in consonant systems; in Hyman, Li; Language, speech and mind. Routledge: London, 62-78.

Lee, S. (2003). The Phonetic Basis of Early Speech Acquisition in Korean. Ph.D. dissertation, The University of Texas at Austin, Austin, TX.

Levelt, W.J.M. (1991). Lexical Access in Speech Production. Blackwell: Cambridge, MA.

Levelt, W.J.M. (1999). Producing spoken language: A blueprint of the speaker, in Brown & Hagoort, The Neurocognition of Language. Oxford University Press, London.

Maddieson, I. (1984). Patterns of sounds. Cambridge University Press: Cambridge, MA.

MacCaffrey, H., Davis, B.L., MacNeilage, P.F. & Von Hapsburg, D (2000). Effects of Multichannel Cochlear Implantation on the Organization of Early Speech. Invited paper, Volta Review, 5-29.

MacNeilage, P.F. & Davis, B.L. (2001). Motor mechanisms in speech ontogeny: Phylogenetic and neural implications. Current Opinion in Neurobiology, 11, 567-569.

MacNeilage, P.F. & Davis, B.L. (2000). Origin of the Internal Structure of Words. Science. 288, 527-531.

MacNeilage, P.F., Davis, B.L., Kinney, A. & Matyear, C.M. (2000a). The motor core of speech: A comparison of serial organization patterns in infants and languages. Invited submission to Special Millenium Issue. Child Development. 71 (1), 153-163.

MacNeilage, P.F., Davis, B.L., Matyear, C.M. & Kinney, A. (2000b). Origin of Speech Output Complexity in Infants and in Languages. Psychological Science. 10 (5) 459-460.

MacNeilage P. (1998), The Frame/Content Theory of Evolution of Speech Production, Behavioral and Brain Sciences, 21, 499-511.

MacNeilage, P.F. & Davis, B.L. (1993). A motor learning perspective on speech and babbling. In Boysson-Bardies,B., Schoen,S., Jusczyk, P., MacNeilage, P. & Morton, J. (eds.) Changes in speech and face processing in infancy: A glimpse at developmental mechanisms of cognition. Dordredcht: Kluwer, 341-352.

MacNeilage, P.F. & Davis, B.L. (1990). Acquisition of speech production: Achievement of segmental independence. In Hardcastle, W.I. & Marchal A. (eds.) Speech Production and Speech Modeling, Dordrecht: Kluwer, 55-68.

Maddieson, I.; Precoda, K. (1992). Syllable structure and phonetic models. UCLA Working Papers Phonet. **78**: 38-49.

Matyear, C. MacNeilage, P.F. & Davis B.L. (1997). Nasalization of vowels in nasal environments in babbling: Evidence for frame dominance. Phonetica. 55, 1-17.

Mitchell, P.R.; Kent, R.D. (1990). Phonetic variation in multi-syllabic babbling. J. Infant Lang. 17, 247-265.

Oller, D.K (1980). The emergence of the sounds of speech in infancy in Yeni-Komshian, Kavanagh, Ferguson, Infant phonology, V 1: Production. Academic Press: New York, 29-42.

Oller, D. K., & Eilers, R. E. (1988). The role of audition in infant babbling. Child Development, *59*(2), 441-449.

Oller, D. K., Eilers, R. E., Bull, D. H., & Carney, A. E. (1985). Prespeech vocalizations of a deaf infant: A comparison with normal metaphonological development. Journal of Speech, & Hearing Research, *28*(1), 47-63.

Oller, D.K. (2000). The Emergence of the Speech Capacity. Erlbaum: London.

Quinlan, P.T. (1992). Oxford psycholinguistic database. Oxford University Press: Oxford.

Redican, W.K. (1975). Facial expressions in nonhuman primates. In L.A. Rosenblum (Ed.) Primate behavior: Developments in field and laboratory research, 4, Academic Press: New York.

Rizzolatti, G., Fadiga L., Gallese, V., and Fogassi, L. (1996) Premotor cortex and the recognition of motor actions. Cogn Brain Res., 3: 131-141.

Rousset, I. From lexical to syllabic organization: Favored and disfavored co-occurrences, Proc. 16th Int. Cong. Phonet. Sci

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

Thorndike, E.; Lorge, I. (1944).The teacher's book of 30,000 words. Bureau of Publications, Teachers College, Columbia University: New York.

Teixeira, E. R. & Davis, B.L. (2003). Early Sound Patterns in the Speech of Two Brazilian Portuguese Speakers, Language & Speech, 45 (2) 179-204.

Van Hooff, J.A.R.A.M. (1962). Facial expressions in higher primates. Symp. Zool. Soc. Lon., 8,97-125.

Van Hooff, J.A.R.A.M. (1967). Facial displays of the catarrhine monkeys and apes. In D. Morris (Ed.) Primate ethology. Weidenfield and Nicholson: London.

Varela, F.J., Thompson, E., Rosch, E. (1997). The Embodied Mind: Cognitive Science and Human Experience, MIT Press, Cambridge, MA.

Vihman, M.M. (1997). Phonological development: the origins of language in the infant. Blackwell: Oxford.

Zlatic, L. MacNeilage, P.F., Matyear, C.L. & Davis, B.L. (1997). Babbling of twins in a bilingual environment. Applied Psycholinguistics 18(4) 453-469.

Zmarich, C.; Lanni, R. A. (1999). Phonetic and acoustic study of babbling in an Italian infant. pp.1103-1104. Proc. 5th Int. Conf. on Spoken Language Processing, Sydney, Australia.