BEHAVIORAL AND BRAIN SCIENCES (1998) **21,** 499–546
*Printed in the United States of America*

# The frame/content theory of evolution of speech production

**Peter F. MacNeilage**
*Department of Psychology,University of Texas at Austin, Austin, TX 78712* **Electronic mail:** *macneilage@mail.utexas.edu*

**Abstract:** The species-specific organizational property of speech is a continual mouth open-close alternation, the two phases of which are subject to continual articulatory modulation. The cycle constitutes the syllable, and the open and closed phases are segments – vowels and consonants, respectively. The fact that segmental serial ordering errors in normal adults obey syllable structure constraints suggests that syllabic "frames" and segmental "content" elements are separately controlled in the speech production process. The frames may derive from cycles of mandibular oscillation present in humans from babbling onset, which are responsible for the open-close alternation. These communication-related frames perhaps first evolved when the ingestion-related cyclicities of mandibular oscillation (associated with mastication [chewing] sucking and licking) took on communicative significance as lipsmacks, tonguesmacks, and teeth chatters – displays that are prominent in many nonhuman primates. The new role of Broca's area and its surround in human vocal communication may have derived from its evolutionary history as the main cortical center for the control of ingestive processes. The frame and content components of speech may have subsequently evolved separate realizations within two general purpose primate motor control systems: (1) a motivation-related medial "intrinsic" system, including anterior cingulate cortex and the supplementary motor area, for self-generated behavior, formerly responsible for ancestral vocalization control and now also responsible for frames, and (2) a lateral "extrinsic" system, including Broca's area and surround, and Wernicke's area, specialized for response to external input (and therefore the emergent vocal learning capacity) and more responsible for content.

**Keywords:** Broca's aphasia; chewing; consonants; lipsmacks; speech evolution syllables; supplementary motor area; vowels; Wernicke's aphasia

## 1. Introduction

This target article is concerned with the evolution of speech production as *action*. The question is, how did we evolve the capacity to *do* what we do with the speech production apparatus when we speak? There will be little concern with the evolution of the conceptual structure that underlies speech actions. Instead, the focus will be on a capability typically taken for granted in current linguistic theory and cognitive science: How do we explain our remarkable capacity for making the serially organized complexes of *movements* that constitute speech?

The basic thesis is quite simple. Human speech differs from vocal communication of other mammals in that we alone superimpose a continual rhythmic alternation between an open and closed mouth (a *frame)* on the sound production process. The likelihood that this cyclicity, associated with the syllable, evolved from ingestive cyclicities (e.g., chewing) is indicated by the fact that much of the new development of the brain for speech purposes occurred in and around Broca's area, in a frontal perisylvian region basic to the control of ingestive movements in mammals. An evolutionary route from ingestive cyclicities to speech is suggested by the existence of a putative intermediate form present in many other higher primates, namely, visuofacial communicative cyclicities such as lipsmacks, tonguesmacks, and teeth chatters. The modification of the frontal perisylvian region leading to syllable production presumably made its other ingestion-related capabilities available for use in modulation of the basic cycle in the form of different consonants and vowels (*content)*. More generally, it is suggested that the control of speech production evolved by descent with modification within two general purpose primate cortical motor control systems, a medial system, associated with vocalization control in all primates, and a lateral system, including Broca's area, that has the necessary emergent vocal learning capacity.

In Darwin's words, evolution is a matter of "descent with modification" (Darwin 1859, p. 420). We must therefore accept the constraint noted by Huxley: "The doctrine of continuity is too well established for it to be permissible to me to suppose that any complex natural phenomenon comes into existence suddenly, and without being preceded by simpler modifications" (Huxley 1917, p. 236). Consequently, the most successful theory of evolution of speech



Peter MacNeilage is Professor of Psychology at the University of Texas at Austin. He has written over 80 papers on the topic of complex action systems and their evolution, including a paper on primate handedness published in *Behavioral and Brain Sciences* (1987, 10:247–63). He is a Fellow of the American Association for the Advancement of Science, the Acoustical Society of America, and the Center for Advanced Study in the Behavioral Sciences. His current research focus is on the acquisition of speech production.

© 1998 Cambridge University Press 0140-525X/98 $12.50

https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

as the action component of language will be the one that best characterizes this descent with modification, with an accurate and dispassionate assessment of prior states and the end state, and of the nature of the difference between them. The best characterization will not be the one that humans often find congenial – one that exults in the glories of the end state and trivializes the precursors. As Darwin (1871) said, "man bears the indelible stamp of his lowly origins" (p. 597).

This characterization immediately rules out any explanation of the ultimate causes of language in terms of the Chomskyan concept of "universal grammar" (Chomsky 1986). This concept is in the tradition of Platonic essentialism (see Mayr 1982, pp. 37–38, on essentialism in biology, and Lakoff 1987, for a characterization of the essentialistic assumptions underlying generative grammar), according to which form has a priori status. In response to the currently accepted view, derived from evolutionary theory, that language has not always been present, Chomsky has departed from both Platonism and orthodox evolutionary theory in implying an instantaneous onset for language form, resulting from "a mutation" (Chomsky 1988, p. 170). However, despite this accommodation to the fact of evolution, there is apparently no room for a role of modification in the Chomskyan scenario.

The following assumptions will be made in the attempt to characterize the state prior to language evolution in this target article: (1) Because the vocal characteristics of call systems of all living nonhuman primates are basically similar despite considerable differences in the closeness of the relations of the various taxa to forms ancestral to humans, it will be assumed that the call systems of forms ancestral to humans were similar to presently observable ones. (2) Most work on brain organization underlying vocal communication in nonhuman primates has been done on two taxa: rhesus monkeys, which are old world monkeys, and squirrel monkeys, which are new world monkeys. These taxa probably had a common ancestry that was also common to humans, about 40 million years ago. The brain organization underlying call production in these two living taxa seems to be relatively similar (Jürgens 1979a). It will be assumed that this similarity owes a good deal more to properties of ancestral brain organization than to convergent evolution of organization radically differing from ancestral organization. It is therefore also assumed that the brain organization underlying call production in these two taxa is basically similar to that of forms ancestral to humans. It is concluded that in underlying brain organization, as well as in vocal production, the problem of accounting for the evolution of human speech production can be considered, for practical purposes, to be the problem of accounting for the change from characteristics displayed by other *living* primates to characteristics of humans.

## 2. Evolution of primate vocal production: Nature of the human-nonhuman difference

### 2.1. Vocal production systems of other mammals

The three main components of the vocal production system of mammals – the respiratory, phonatory, and articulatory components – are shown schematically in Figure 1. They are shown in the typical horizontal plane characteristic of quadrupeds. With the advent of bipedalism in hominids,

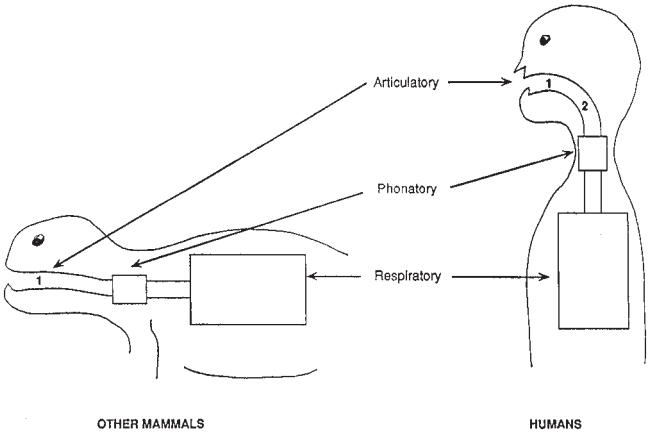



Figure 1. Schematic view of the three main components of the vocal production apparatus.

the respiratory and phonatory components take on a vertical orientation. In addition, as shown in this figure, in advanced hominids the posterior part of the articulatory system takes on a vertical configuration, but the anterior part does not, resulting in a two-tubed vocal tract (perhaps in the last few hundred thousand years according to Lieberman 1984).

The main role of the respiratory component in sound production is to produce an outward flow of air under pressure (Hixon 1973). Phonation (or voicing) is produced when the vocal folds are brought together in such a way that they vibrate when activated by the outward air flow (Negus 1949). The articulatory component – basically the mouth – is usually opened at least once for a vocal episode, and the shape of the cavity between lips and larynx – the vocal tract – modulates the voice source in the form of resonances (Fant 1960). The value of the evolution of the two-tubed vocal tract (Lieberman 1984) in hominids was that it considerably increased the acoustic potential for making different sounds (Carré et al. 1995). However, the question being raised here is: How did humans evolve the *organizational* capacity to make use of this potential by producing rapid and highly variegated sound sequences in syllabic packages?

Except for humans, mammals typically have a very small repertoire of different calls, with some seeming to involve a graded continuum. For example, in a recent study of gelada baboon vocalizations (Aich et al. 1990) "at least 22 acoustically different vocal patterns" were distinguished. Their distinctively *holistic* character, lacking independently variable internal subcomponents, is indicated by the fact that they are often given names with single auditory connotations. Names given to gelada baboon calls by Dunbar and Dunbar (1975) include "moan," "grunt," "vocalized yawn," "vibrato moan," "yelp," "hnn pant," "staccato cough," "snarl," "scream," "aspirated pant," and "how bark." Some calls of other primates occur only alone, some alone and in series, and some only in series. Although it occurs "often" (Marler 1977, p. 24), different acoustic units are not typically combined into series in other primates, and when they are, different arrangements of internal subcomponents do not seem to have separate meanings in themselves (e.g., Robinson 1979).


https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

### 2.2. The nature of speech

The main difference between speech and other mammalian call systems involves the articulatory component. In all mammals, the operation of the respiratory and phonatory components can be most generally described in terms of modulated biphasic cyclicities. In respiration, the basic cycle is the inspiration–expiration alternation and the expiratory phase is modulated to produce vocalizations. In the phonatory system, the basic cycle is the alternation of the vocal folds between an open and closed position during phonation (*voicing* in humans; Broad 1973). This cycle is modulated in its frequency, presumably in all mammals, by changes in vocal fold tension and subglottal pressure level, producing variations in perceived pitch.

The articulatory system in nonhuman mammals is typically only used in an open configuration during call production, although some calls in some animals (e.g., "girneys" in Japanese macaques – see Green 1975) seem to involve a rhythmic series of open–close alternations. However, in human speech in general, the fact that the vocal tract alternates more or less regularly between a relatively open and a relatively closed configuration (open for vowels and closed for consonants) is basic enough to be a defining characteristic (MacNeilage 1991a). With the exception of a few words consisting of a single vowel, *virtually every utterance of every speaker of every one of the world's languages* involves an alternation between open and closed configurations of the vocal tract. As noted earlier, the *syllable,* a universal unit in speech, is defined in terms of a nucleus with a relatively open vocal tract and margins with a relatively closed vocal tract. Modulation of this open-close cycle in humans takes the form of typically producing different basic units – consonants and vowels, collectively termed *phonemes* – in successive closing and opening phases. Thus, human speech is distinguished from other mammalian vocal communication, in movement terms, by the fact that a third, articulatory, level of modulated cyclicity continuously coexists with the two levels present in other mammals.

Figure 2 is a schematic view of the structure of the English word *tomato.* It can be described as consisting of two levels, suprasegmental and segmental. The segmental level, consisting of consonants and vowels, can be further divided into a number of subattributes or features. (In more behaviorally oriented treatments, subattributes of phonemes are described in terms of gestures, e.g., Browman & Goldstein 1986.) For example, for the sound [t], a featural description would be applied to its voicing properties, the place in the vocal tract at which occlusion occurred and the fact that it involves a complete occlusion of the vocal tract. At the suprasegmental level, the term *stress* refers roughly to the amount of energy involved in producing a syllable, which is correlated with its perceptual prominence. In English at least, more stressed syllables tend to be louder and have higher fundamental frequencies and longer durations. *Intonation* refers to the global pattern of fundamental frequency (rate of vocal fold vibration). In multisyllabic words spoken in isolation, and in simple declarative sentences such as "The boy hit the ball," there is a terminal fall in fundamental frequency. The syllable lies at the interface between the suprasegmental and the segmental levels. At the suprasegmental level it is the unit in terms of which stress is distributed, a unit of rhythmic organization, and a point of inflexion for intonation contours. At the segmental level it provides an organizational superstructure for the distribution of consonants and vowels. (For further detail see Levelt 1989, Ch. 8.)

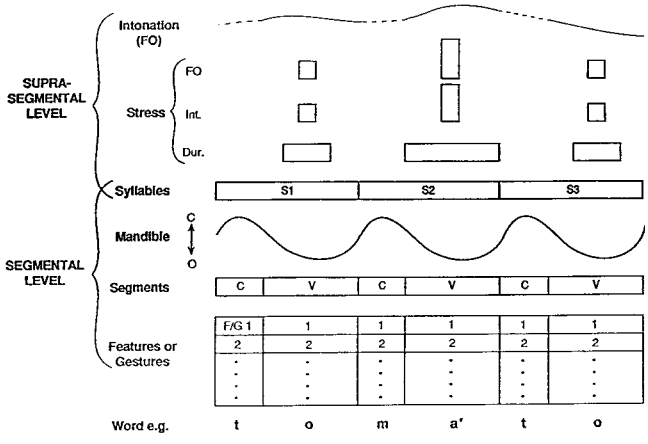



Figure 2. Schematic view of the organization of speech in the word "tomato".

## 3. How is the new human capability organized? In a frame/content mode

### 3.1. Serial ordering errors in speech

How do we discover the organizational principles underlying syllabic frames and their modulation by internal content? Normal speakers sometimes make errors in the serial organization of their utterances. It was Lashley (1951) who realized that serial ordering errors provide important information about both the functional units of action and their serial organization. At the level of sounds (rather than words) the most frequent unit to be misplaced is the single segment (consonant or vowel). For example, in a corpus collected by Shattuck-Hufnagel (1980), approximately about two thirds of the errors involved single segments. The other errors involved for the most part subsyllabic groupings of segments.

There is some agreement on the existence of five types of segmental speech error, often called "exchange" (Spoonerisms), "substitution," "shift," "addition," and "omission" errors. In previous discussions of the implications of speech errors, the author and colleagues have focussed primarily on exchange errors (MacNeilage 1973; 1985; 1987a; 1987b; MacNeilage et al. 1984; 1985) because they are the only relatively frequently occurring type in which the source of the unit can be unequivocally established. However, much evidence from other error types is consistent with that from exchange errors.

The central fact about exchange errors is that in virtually all segmental exchanges, the units move into a position in syllable structure similar to that which they vacated: syllable-initial consonants exchange with other syllable-initial consonants, vowels exchange with vowels, and syllable-final consonants exchange with other syllable-final consonants. For example, Shattuck-Hufnagel (1979) reported that of a total of 211 segmental exchanges between words, "all but 4 take place between phonemes in similar positions in their respective syllables" (p. 307).



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

Examples from Fromkin (1973) are:

Initial consonants: well made – mell wade
Vowels: ad hoc – odd hack
Final consonants: top shelf – toff shelp

This result, which is widely attested in studies of both spontaneous and elicited errors (Levelt 1989) demonstrates that there is a severe syllable position constraint on the serial organization of the sound level of language. Most notably, the position-in-syllable constraint seems virtually absolute in preserving a lack of interaction between consonants and vowels. There are numbers of consonant-vowel and vowel-consonant syllables in English that are mirror images of each other (e.g., eat vs. tea; no vs. own; *ab*stract vs. *ba*stract). Either form therefore naturally occurs as a sequence of the two opposing vocal tract phases, but exchange errors that would turn one such form into the other are not attested.

### 3.2. Metaphors for speech organization: Slot/segment and frame-content

According to Shattuck-Hufnagel (1979), these error patterns imply the existence of a scan-copy mechanism that scans the lexical items of the intended utterance for representation of segments and then copies these representations into slots in a series of canonical syllable structure matrices. The fundamental conception underlying this "slot/segment" hypothesis is that "slots in an utterance are represented in some way during the production process independent of their segmental contents" (Shattuck-Hufnagel 1979, p. 303). It is this conception that also underlies the frame/content (F/C) metaphor used by me and my colleagues (MacNeilage et al. 1984; 1985; MacNeilage 1985; 1987a; 1987b) and by Levelt (1989). The only difference lies in the choice of terms for the two components. In the present terms, *syllable-structure frames* are represented in some way during the production process independent of *segmental content elements.*

The speech errors that reveal the F/C mode of organization of speech production presumably occur at the stage of interfacing the lexicon with the motor system. The motor system is required to both produce the overall rhythmic organization associated with syllables, basically by means of an open-close alternation of the vocal tract, and to continually modulate these cycles by producing particular consonants and vowels during closing and opening phases. Rather than there being holistic chunking of output into an indissoluble motor package for each syllable, there may have developed, *in the production system,* some natural division of labor whereby the basic syllabic cycle and the phasic modulations of the cycle are separately controlled. Thus, perhaps when frame modulation, by means of varying consonants and vowels, evolved as a favored means of increasing the message set, the increasing load on this aspect of production led to the development of a separate mechanism for its motor control.

According to the above conception, which will be amplified in subsequent discussion, fundamental phylogenetic properties of the motor system have played the primary role in determining the F/C structure of speech. It is assumed that as this occurred the consequences of the two-part division of labor then ramified into the organization of the prior stage of lexical storage. There is good evidence that there is, in fact, independent lexical representation of segmental information and information about syllable structure in the mental lexicon. This evidence comes from a set of studies on the "tip of the tongue" (TOT) phenomenon, which occurs when people find themselves able to retrieve some information about the word they wish to produce but cannot produce the whole word. Levelt (1989) concludes that "lexical form information is not all-or-none. A word's representation in memory consists of components that are relatively accessible and there can be metrical information about the number and accents of syllables without these syllables being available" (p. 321).

The conception of the syllable as the *receptacle* for segments during motor organization is supported by another body of evidence. Garrett (1988) has pointed out that there is little evidence that syllables themselves are moved around in serial ordering errors "except where the latter are ambiguous as to their classification (i.e., they coincide with morphemes, or the segmental makeup of the error unit is ambiguous)" (p. 82). Thus, "syllables appear to constrain error rather than indulge in it." (For a similar conclusion, see Levelt 1989, p. 322.)

### 3.3. Lack of evidence for subsegmental units

It is of interest to note that in emphasizing this dual-component (syllable and segment) conception of speech production, no role is accorded to the most nested subcomponent in the linguistic description of syllable structure, the distinctive feature, or its functional counterpart, the gesture, the units most favored in current phonologic and phonetic conceptions of the organization of speech. This contrarian stance is taken primarily on the grounds of the paucity of evidence from speech errors that the feature/gesture is an independent variable in the control of speech production. The fact that members of most pairs of segments involved in errors are similar, differing only by one feature, sometimes has been taken to mean that the feature is a functional unit in the control process. However, the proposition that phonetic similarity is a variable in potentiating errors of serial organization can be made without dependence on an analysis in terms of features. When two exchanged segments differ by one feature, it cannot be determined whether features or whole segments have been exchanged; but as Shattuck-Hufnagel and Klatt (1979) have pointed out, when the two segments participating in an exchange error differ by more than one feature, a parsimonious interpretation of the view that features are functional units would suggest that the usual number of features that would be exchanged would be one. However, in an analysis of 72 exchange errors in which the members of the pairs of participating segments differed by more than one feature, there were only three cases where only a single feature was involved in the exchange. Of course, this is not conclusive evidence against the independence of features/gestures as units in the control process, but it does serve to encourage a conception of production in which their independence is not required.

### 3.4. Speech and typing

A perspective on this dual-component view of speech production organization can be gained by comparing it with another language output behavior: typing. There is evidence to suggest that there is a considerable commonality be-


https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

tween spoken language and typing – even copy typing – in *early* stages of the process of phonological output, stages in which there is a role of the lexicon. For example, Grudin (1981) found that on 11 of 15 occasions, copy typists spontaneously corrected the spelling of a misspelled word with which they were inadvertently presented. However, typing does not possess an F/C mode of organization. Any typist knows that, in contrast with spoken language, exchange errors occur not between units with comparable positions in an independently specified superordinate frame structure, but simply between adjacent letters (MacNeilage 1964). This is true whether the units are in the same syllable or in different syllables. In addition, unlike in speech, there is no constraint against exchanging actions symbolizing consonants and actions symbolizing vowels. Vowel and consonant letters exchange with each other about as often as would be predicted from the relative frequency with which vowel letters and consonant letters appear in written language (MacNeilage 1985). Nespoulous et al. (1985) have reported a similar freedom from phonotactic constraints of the language in agraphics.

In concluding this section on adult speech organization, it should be emphasized that the present focus on the F/C dichotomy is not simply a case of deification of some marginal phenomenon. As Levelt puts it: "Probably the most fundamental insight from modern speech error research is that a word's skeleton or frame and its segmental content are independently generated" (1992, p. 10). Speech error data have in turn been the most important source of information in the psycholinguistic study of language production.

## 4. How did the frame/content mode evolve?

### 4.1. Evolution as tinkering

François Jacob's metaphor of evolution as tinkering has gained wide acceptance (Jacob 1977). Evolution does not build new structures from scratch as an engineer does. Instead it takes whatever is available, and, where called for by natural selection, molds it to new use. This is presumably equally true for structures and behaviors. Of course, there are plenty of examples of this in the evolution of vocalization. No structure in the speech production system initially evolved for vocalization. Our task is to determine what modifications of existing capacities led to speech. Specifically, the question is: How was the new articulatory level of modulated cyclicity tinkered into use?

### 4.2. Cyclicities and tinkering

An obvious answer suggests itself. The oral system has an extremely long history of ingestive cyclicities involving mandibular oscillation, probably extending back to the evolution of the first mammals, circa 200 million years ago. Chewing, licking and sucking are extremely widespread mammalian activities, which, in terms of casual observation, have obvious similarities with speech, in that they involve successive cycles of mandibular oscillation. If ingestion-related mandibular oscillation was modified for speech purposes, the articulatory level would be similar to the other two levels in making use of preexisting cyclicities. The respiratory cycle originally evolved for gas exchange, and the larynx initially evolved as a valve protecting the lungs from invasion by fluids. Presumably, vocal fold cyclicities were initially adventitious results of release of air through the valve under pressure, a phenomenon similar to that sometimes observed in the anal passage, but one that presumably had more potential for control.

It is well known that biphasic cycles are the main method by which the animal kingdom does work that is extended in the time domain. Many years ago, Lashley (1951) attempted, more or less unsuccessfully, to bring to our attention the importance of rhythm generators as a basis for serially organized behaviors, even behaviors as complex as speech. Examples of such biphasic cycles are legion: locomotion of many different kinds in aquatic, terrestrial, and aerial media, heartbeat, respiration, scratching, digging, copulating, vomiting, milking cows, pedal alarm "calling" in rabbits, cyclical ingestive processes, and so forth. The conservative connotation of the tinkering metaphor is applicable to the fact that biphasic cyclicities, once invented, do not appear to be abandoned but are often modified for uses somewhat different than the original one. For example, Cohen (1988) makes the astonishing claim that an evolutionary continuity in a biphasic vertebrate locomotory cycle of flexion and extension can be traced back over a period of one half billion years: "There is . . . a clear phylogenetic pathway from lampreys to mammalian quadrupeds for the locomotor central pattern generator (CPG)" (p. 160). She points out that "With the evolution of more sophisticated and versatile vertebrates, more levels of control have been added to an increasingly more sensitive and labile CPG coordinating system." She concludes, however, that "In this view the basic locomotor CPG need change very little to accommodate the increasing demands natural selection placed on it" (p. 161).

### 4.3. Ingestive cyclicities

Ingestive oral cyclicities are similar to locomotion in that they have a CPG in the brainstem that has similar characteristics across a wide range of mammals. In fact, the similarity between the locomotor and ingestive CPGs is sufficiently great that Rossignol et al. (1988) were motivated to suggest a single neural network model for these two CPGs and the CPG for respiration. Lund and Enomoto (1988) characterize mastication as "one of the types of rhythmical movements that are [sic] made by coordinated action of masticatory, facial, lingual, neck and supra- and infra-hyoid muscles" (p. 49). In fact, this description is apt for speech. The question is whether speech would develop an entirely new rhythm generator, with its own totally new superordinate control structures, which could respond to coordinative demands similar to those made on the older system, if evolution is correctly characterized as a tinkering operation, making conservative use of existing CPGs. The answer to this question must be *No*! If so, then it is not unreasonable to conclude that speech makes use of the same brainstem pattern generator that ingestive cyclicities do, and that its control structures for speech purposes are, in part at least, shared with those of ingestion.

In coming to this conclusion one needs to resist a tendency to regard mastication as too simple to be a candidate for tinkering into speech. As Luschei and Goldberg (1981) point out, mastication is "a rhythmic activity that seems to proceed successfully in a highly 'automatic' fashion, even in the face of wide variation in the loads presented by eating



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

different food materials" (p. 1237). However, they warn us that "movements of mastication are actually quite complex and they must bring the teeth to bear on the food material in a precise way" (p. 1238). In addition, they note that " . . . the mandible is often used in a controlled manner for a variety of tasks. For the quadrupeds, in particular, the mandible constitutes an important system for manipulation of objects in the environment" (p. 1238). The inaccessability of the masticatory system to direct observation presumably contributes to a tendency to underestimate its prowess. The reader may have shared the author's surprise, on biting his tongue, that it does not occur more often.

Perhaps part of the reason that so little attention has been given to the possibility that ingestive cyclicities were precursors to speech is that speech is a quite different function from ingestion. However, functional changes that occur when locomotor cyclicities of the limbs are modified for scratching and digging do not prompt a denial of the relation of these functions to locomotion. In my opinion, it is the anthropocentric view of speech as having exalted status that is the main reason for the neglect of the possibility that actions basic to it may have had ingestive precursors.

### *4.4. Visuofacial communicative cyclicities*

If the articulatory cyclicity of speech indeed evolved from ingestive cyclicities, how would this have occurred? An important fact in this regard is that mandibular cyclicities, though not common in nonhuman vocalization systems, are extremely common as faciovisual communicative gestures. "Lipsmacks," "tonguesmacks," and "teeth chatters" can be distinguished. Redican (1975) describes the most common of these, the lipsmack, as follows: "The lower jaw moves up and down but the teeth do not meet. At the same time the lips open and close slightly and the tongue is brought forward and back between the teeth so that the movements are usually quite audible. . . . The tongue movements are often difficult to see, as the tongue rarely protrudes far beyond the lips" (p. 138). Perhaps these communicative events evolved from ingestive cyclicities.

It is surprising that more attention has not been drawn to the similarity between the movement dynamics of the lipsmack and the dynamics of the syllable (MacNeilage 1986). The up and down movements of the mandible are typically reduplicated in a rhythmic fashion in the lipsmack, as they are in syllables. In addition to its similarity to syllable production in motor terms, there are a number of other reasons to believe that the lipsmack could be a precursor to speech. First, it is analogous to speech in its ubiquity of occurrence. Redican (1975) believes that it may occur in a wider variety of social circumstances than any of the other facial expressions that he reviewed. A second similarity between the lipsmack and speech is that both typically occur in the context of positive social interactions. A third similarity is that, unlike many vocal calls of the other primates, the lipsmack is an accompaniment of one-on-one social interactions involving eye contact, and sometimes what appears to be turn-taking. This is the most likely context for the origin of true language.

Finally, in some circumstances the lipsmack is accompanied by phonation. Andrew (1976) identifies a class of "humanoid grunts" involving low frequency phonation in baboons, sometimes combined with lipsmacking. In the case he studied most intensively, mandibular lowering was accompanied by tongue protrusion, and mandibular elevation by tongue retraction. Green (1975) describes a category of "atonal girneys" in which phonation is modulated "by rapid tongue flickings and lipsmacks." Green particularly emphasizes the labile morphology of these events, stating that "a slightly new vocal tract configuration may be assumed after each articulation" (p. 45). Both Andrew and Green suggest that these vocal events could be precursors to speech.

Exactly how might ingestive cyclicities get into the communicative repertoire? Lipsmacks occurring during grooming often have been linked with the oral actions of ingestion of various materials discovered during the grooming process, because they often precede the ingestion of such materials. In young infants they have been characterized as consisting of, or deriving from, nonnutritive sucking movements. It does not seem too far fetched to suggest that gestures anticipatory to ingestion may have become incorporated into communicative repertoires.

## 5. Phylogeny and ontogeny: Development of the frame/content mode

### *5.1 Manual ontogeny recapitulates phylogeny*

The claim, originating with Haeckel (1896), that ontogeny recapitulates phylogeny, has been discredited in a number of domains of inquiry (Gould 1977; Medicus 1992). However, in the realm of human motor function there is some evidence in favor of it. Paleontological evidence, plus the existence of living forms homologous with ancestral forms, allows a relatively straightforward reconstruction of the general outlines of the evolutionary history of the hand (Napier 1962). Mammals ancestral to primates are considered to have the property of convergence-divergence of the claws or paws of the forelimbs but not to have prehensility (the capability of enclosing an object within the limb extremity). This is considered to have first developed with the hand itself in ancestral primates (prosimians) about 60 million years ago. Precise control of individual fingers, including opposability of the thumb, which allows a precision grip, only became widespread in higher primates, whose ancestral forms evolved about 40 million years ago (MacNeilage 1989). In human infants, while convergence-divergence is present from birth, spontaneous manual prehension does not develop until about 3 to 4 months of age (Hofsten 1984), and "it is not until 9 months of age that infants start to be able to control relatively independent finger movements" (Hofsten 1986).

### *5.2. Speech ontogeny: Frames, then content*

A similar relationship exists between the putative phylogeny of speech and its ontogeny. Infants are born with the ability to phonate, which involves the cooperation between the respiratory and phonatory systems characteristic of all mammals. Meier et al. (1997) have recently found that infants may produce "jaw wags," rhythmic multicycle episodes of mouth open-close alternation without phonation – a phenomenon similar to lipsmacks – as early as 5 months of age. Then, at approximately 7 months of age, infants begin to babble, producing rhythmic mouth open-close alternations accompanied by phonation.

Work with Davis and other colleagues has shown convincingly that the main source of variance in the articula-


https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

tory component of babbling (7–12 months) and subsequent early speech (12–18 months) is mandibular oscillation. The ability of the other articulators – lips, tongue, soft palate – to actively vary their position from segment to segment, and even from syllable to syllable, is extremely limited. We have termed this phenomenon *frame dominance* (Davis & MacNeilage 1995).

We have hypothesized that frame dominance is indicated by five aspects of babbling and early speech patterns. Three of these hypotheses involve relations between consonants and vowels in consonant-vowel syllables, the most favored syllable type in babbling and early speech, and the other two involve relations between syllables. The first two hypotheses concern the possible lack of independence of the tongue within consonant-vowel syllables: (1) Consonants made with a constriction in the front of the mouth (e.g., "d," "n") will be preferentially associated with front vowels. (2) Consonants made with a constriction in the back of the mouth (e.g., "g") will be preferentially associated with back vowels. (3) A third hypothesis is that consonants made with the lips (e.g., "b," "m") will be associated with central vowels; that is, vowels that are neither front nor back. It was suggested that, because no direct mechanical linkage could be responsible for lip closure co-occurring with central tongue position, these syllables may be produced simply by mandibular oscillation, with both lips and tongue in resting positions. These consonant-vowel syllable types were called *pure frames.*

The lack of independent control of articulators other than the mandible during the basic oscillatory sequence of babbling is further illustrated by the fact that, approximately 50% of the time, a given syllable will be followed by the same syllable (Davis & MacNeilage 1995). This phenomenon has been called *reduplicated babbling,* and apparently involves an unchanging configuration of the tongue, lips, and soft palate from syllable to syllable. It was further hypothesized that even when successive syllables differ, (a phenomenon called *variegated babbling)* the difference might most often be related to frame control, reflected in changes in the elevation of the mandible between syllables. In general it was proposed that changes in the vertical dimension, which could be related to the amount of elevation of the mandible, would be more frequent than changes in the horizontal dimension. Changes in the horizontal dimension would be between a lip and tongue articulation for consonants, or changes in the front-back dimension of tongue position for consonants or for vowels. The resultant hypotheses were: (4) There will be relatively more intersyllabic changes in manner of articulation (specifically, amount of vocal tract constriction) than in place of constriction for consonants. (5) There will be relatively more intersyllabic changes in tongue height than in the front-back dimension for vowels.

To date, in three papers ( Davis & MacNeilage 1995; MacNeilage & Davis 1996; Zlatic et al. 1997) we have reported a total of 99 tests in 14 infants of these five hypotheses regarding the predominant role of frames in prespeech babbling, early speech, and babbling concurrent with early speech. Of these 99 tests, 91 showed positive results, typically at statistically significant levels, 6 showed countertrends, and 2 showed an absence of trend.

Is it a mere coincidence that the frame dominance pattern that we have found in both babbling and the earliest words is similar to the pattern postulated here for the earliest speech of hominids, or is this pattern showing us the most basic properties of hominid speech production? If the earliest speech patterns were not like this, what *were* they like and why? And why has this question not received attention?

Another way of looking at this matter is to argue that modern hominids have evolved higher levels of both manual and vocal skills than their ancestors, but that this skill only becomes manifest later in development. The question of skill development in speech production requires some background. Most work on the sound preferences in babbling and early words has been done on consonants. Labial, alveolar, and velar stops (e.g., "b," "d," and "g," respectively) and labial and alveolar nasals ("m," "n") are most favored. Lindblom and Maddieson (1988) have classified consonants into three levels of difficulty, in terms of the number of separate action subcomponents they require. Ordinary stops and nasals are in the "simple" category. In fact, even though within the simple category, consonants that are widely considered to be more difficult to produce than ordinary stops and nasals (e.g., liquids, such as those written in English orthography as "r" and "l," and fricatives such "th") are relatively infrequent in babbling and early words (Locke 1983), and even remain problematic for life for some speakers. Thus, the progression in development of consonant production is from simple sounds to those that can be considered to require more skill.

The possibility that this was also the sequence of events in the evolution of language is supported by another aspect of the work of Lindblom and Maddieson (1988). In a survey of the consonant inventories of languages, they found that languages with small inventories tended to have only their "simple" consonants, languages with medium-sized inventories differed mainly by also including "complex" consonants, and languages with the largest inventories tended to also add "elaborated" consonants, the most complex subgroup in the classification. Presumably, the first true language(s) had a small number of consonants. It seems that the only way that the beyond-chance allocation of difficult consonants to languages with larger inventories can be explained is by arguing that they tended to employ consonants of greater complexity as the size of their inventories increased. If so, the tendency for infants to add more difficult consonants later in acquisition suggests that ontogeny recapitulates phylogeny.

### 5.3. Sound pattern of the first language

If babbling and early speech patterns are similar to those of the first language, what was it like? I have proposed "that the conjoint set of sounds and sound patterns favored in babbling *and* in the world's languages constitutes, in effect the fossil record of true speech" (MacNeilage 1994). The proposed consonants are the voiceless unaspirated stops [p], [t], and [k] (as in "bill," "dill," "gill") and the nasals [m] and [n] (as in "man"). (The brackets denote phonetic symbols.) The two semivowels [w] and [j] (as in "wet" and "yet") can also occupy the consonant position in syllables. The three vowels are versions of the three point vowels [i], [u], and [a]. Only the consonant-vowel syllable type is allowed, either alone or with one reiteration. Some constraints on possible intersyllabic combinations, similar to those observed in babbling and early speech, are imposed. An initial corpus of 102 words is proposed.



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

### 5.4. Frames and rhythmic behavior

Phylogeny can profitably be characterized as a succession of ontogenies. The important role in evolution of biphasic cycles with their basically fixed rhythms is paralleled by their important role in ontogeny. From the beginning of babbling, utterances typically have a fixed rhythm in which the syllable frame is the unit. Mastery of rhythm does not develop from nonrhythmicity as it does in learning to play the piano. I appeal to the intuition of the reader as parent or supermarket shopper that intersyllable durations of babbling utterances often sound completely regular.

This initial rhythmicity provides a basis for the control of speech throughout life. For example, Kozhevnikov and Chistovich (1965) have observed that when speakers changed speaking rate the relative duration of stressed and unstressed syllables remained more or less constant, suggesting the presence of a superordinate rhythmic control generator related to syllable structure. They also noted that the typical finding of shorter segment durations in syllables with more segments reflected an adjustment of a segmental component to a syllabic one.

Thelen (1981) has emphasized the fact that babbling is simply one of a wide variety of repetitive rhythmic movements characteristic of infants in the first few months of life: "kicking, rocking, waving, bouncing, banging, rubbing, scratching, swaying . . . " (p. 238). As she notes, the behavior "stands out not only for its frequency but also for the peculiar exuberance and seemingly pleasurable absorption often seen in infants moving in this manner" (p. 238). She believes that such "rhythmic stereotypies are transition behavior between uncoordinated behavior and complex, coordinated motor control." In her opinion, they are "phylogenetically available to the immature infant. In this view, rhythmical patterning *originating as motor programs essential for movement control* . . . [emphasis mine] are 'called forth,' so to speak, during the long period before full voluntary control develops, to serve adaptive needs later met by goal-corrected behavior" (p. 253). She suggests an adaptive function for such stereotypies, as aids to the infants in becoming active participants in their social environment. This, in turn, suggests a scenario whereby the child could have become father to the man so to speak, in the evolution of speech, by encouraging use of rhythmic syllabic vocalization for adult communication purposes. (See also Wolff 1967; 1968, for an earlier discussion of a similar thesis.)

### 5.5. Perceptual consequences of the open-close alternation

The focus of this target article is speech *production.* From this standpoint, the evolution of the mouth open-close alternation for speech is seen as the tinkering of an already available motor cyclicity into use as a general purpose carrier wave for time-extended message production, with its subsequent modulation increasing message set size. However, it has also been pointed out that the open-close alternation confers perceptual benefits. In particular, the acoustic transients, which are associated with consonants and accompany onset and offset of vocal tract constriction, are considered to be especially salient to the auditory system (e.g., Stevens 1989). The ability to produce varied transients at high rates may have been an important hominid-specific communicative development. In addition, the regularly repeating high amplitude events provided by the vowels may have played an important role in inducing rhythmic imitations.

## 6. Comparative neurobiology of the frame/content mode

### 6.1. The evolution of Broca's area

The possibility that the mandibular cycle is the main articulatory building block of speech gains force from the fact that the region of the inferior frontal lobe that contains Broca's area in humans is the main cortical locus for the control of ingestive processes in mammals (Woolsey 1958). In particular the equivalents in the monkey of Brodmann's area 44 – the posterior part of classical Broca's area – and the immediately posterior area 6 have been clearly implicated in mastication (Luschei & Goldberg 1981), and electrical stimulation of area 6 in humans evokes chewing movements (Foerster 1936a). In addition in recent high resolution positron emission tomography (PET) studies, cortical tissue at the confluence of areas 44 and 6 has been shown to be activated during speech production. Figure 3 shows regions of activation of posterior inferior frontal cortex in two studies in which subjects spoke written words (Petersen et al. 1988 [square]; LeBlanc 1992 [circle]). The points are plotted on horizontal slice $z = 16$ mm of the normalized human brain coordinates made available by Talairach (Talairach & Tornoux 1988). The figure was generated by use of the Brainmap database (Fox et al. 1995) Both areas straddle the boundary between Brodmann's areas 6 and 44. Fox (1995) reports additional evidence of joint ac-

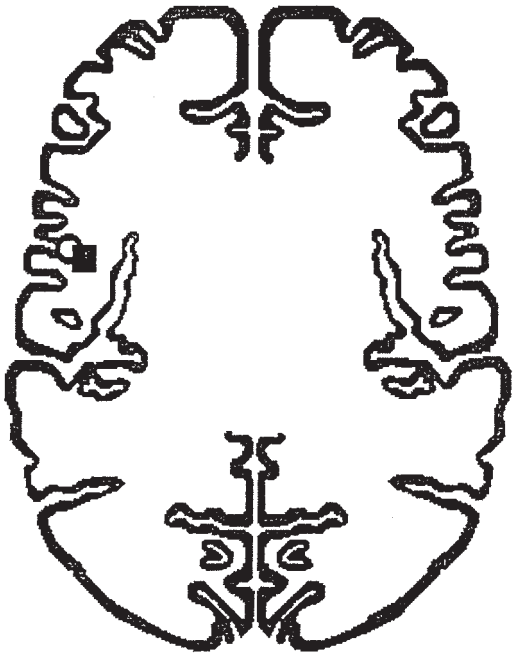


Figure 3. Sites in lateral premotor cortex significantly active during spoken word response to auditory and visual stimulation in the positron emission tomography studies of Peterson et al. (1988) and LeBlanc (1992).


https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

tivation of areas 6 and 44 during single word speech.

Of course, a landmark event in the history of neuroscience was the discovery that Broca's area plays an important role in the motor control of speech. More recently a good deal of significance has been attached to the discovery by paleontologists that the surface configuration of the cortex in this region underwent relatively sudden changes in *Homo habilis* (e.g., Tobias 1987). The question of exactly why it was this particular area of the brain that took on this momentous new role has received little attention. Perhaps part of the answer may come not only from the recognition of the importance of our ingestive heritage in the evolution of speech, but also when one acknowledges the more general fact that the main change from other primate vocalization to human speech has come in the articulatory system. Consistent with this fact, bilateral damage to Broca's area and the surrounding region does not interfere in any obvious way with monkey vocalization (Jürgens et al. 1982), but unilateral damage to the region of Broca's area in the left hemisphere, if sufficiently extensive, results in a severe deficit in speech production. However, despite the involvement of Broca's area in the control of the articulatory apparatus, caution is advised in drawing implications from this part of *Homo habilis* morphology for the evolution of speech. This region is also involved in manual function in monkeys (Gentilucci et al. 1988; Rizzolatti et al. 1988) and in humans (Fox 1995).

### 6.2. Medial frontal cortex and speech evolution

At first glance, evolution of a new vocal communication capacity in Broca's area of humans appears to constitute a counterexample to Darwin's basic tenet of descent with modification. It has often been considered to be an entirely new development (e.g., Lancaster 1973; Myers 1976; Robinson 1976). The main region of cortex controlling vocal communication in monkeys is anterior cingulate cortex, on the medial surface of the hemisphere (Jürgens 1987). Vocalization can be evoked by electrical stimulation of this region and damage to it impairs the monkey's ability to voluntarily produce calls on demand (e.g., in a conditioning situation). However, a clue to the evolutionary sequence of events for speech comes from consideration of the supplementary motor area (SMA) an area immediately superior to anterior cingulate cortex and closely connected with it. While this area has not been implicated in vocal communication in monkeys, it is consistently activated in brain imaging studies of speech (Roland 1993) and it is active even when the subjects merely think about making movements (Orgogozo & Larson 1979). It was given equal status with Broca's and Wernicke's areas as a language area in the classic monograph of Penfield and Roberts (1959).

Two properties of the SMA are of particular interest in the context of the F/C theory. A number of investigators have reported that electrical stimulation of this area often makes patients involuntarily produce simple consonant-vowel syllable sequences such as "dadadada" or "tetetete" (Brickner 1940; Chauvel 1976; Dinner & Luders 1995; Erikson & Woolsey 1951; Penfield & Jasper 1954; Penfield & Welch 1951; Woolsey et al. 1979). Penfield and Welch concluded from their observations of rhythmic vocalizations that "these mechanisms, which we have activated by gross artificial stimuli, may, however, under different conditions, be important in the production of the varied sounds which men often use to communicate ideas" (p. 303). I believe that this conclusion was of profound importance for the understanding of the mechanism of speech production and its evolution, but apparently it has been totally ignored.

In addition, Jonas (1981) has summarized eight studies of irritative lesions of the SMA that have reported involuntary production of similar sequences by 20 patients. The convergence of these two types of evidence strongly suggests that the SMA is involved in frame generation in modern humans.

It thus appears that the evolution of a communicative role for Broca's area was not an entirely de novo development. It is more likely that when mandibular oscillations became important for communication, their control for this purpose shifted to the region of the brain that was already most important for control of communicative output – medial cortex. However, it may have been that, once the mandibular cycle was co-opted for communicative purposes, the overall motor abilities associated with ingestion also became available for tinkering into use for communicative purposes. This is consistent with the fact that a typical result of damage to Broca's area is what has been called "apraxia of speech" – a disorder of motor programming revealed by phonemic paraphasias and distortions of speech sounds (e.g., MacNeilage 1982).

### 6.3. Medial and lateral premotor systems

Further understanding of this particular distribution of speech motor roles and how they relate to properties of manual control can be gained by viewing the *overall* problem of primate motor control from a broader perspective. It is now generally accepted that the SMA and inferior premotor cortex of areas 6 and 44 are the main areas of premotor cortex for two fundamentally different motor subsystems for bodily action in general (e.g., Eccles 1982; Rizzolatti et al. 1983; Goldberg 1985; 1992; Passingham 1987). Using the terminology of Goldberg, anterior cingulate cortex and the SMA are part of a medial premotor system (MPS), associated primarily with intrinsic, or self-generated, activity, while the areas of inferior premotor cortex are part of a lateral premotor system (LPS), associated primarily with "extrinsic" actions; that is, actions responsive to external stimulation. The connectivity of these two premotor areas is consistent with this proposed division of labor. While the sensory input to the SMA is primarily from deep somatic afferents, inferior premotor cortex receives heavy multimodal sensory input – somatic input from anterior parietal cortex, visual input primarily from posterior parietal cortex, and auditory input from superior temporal cortex, including Wernicke's area in the left hemisphere of humans (Pandya 1987).

This basic action dichotomy has been well established by studies involving both lesion and electrical recording in monkeys. The MPS has been shown to be primarily involved in tasks in which monkeys produce sequences of previously learned manual actions with no external prompting, while LPS is primarily involved in sequencing tasks in which the component acts are cued by sensory stimulation (e.g., lights) (Tanji et al. 1995). The human equivalent of the findings from monkey lesion studies of MPS is an initial akinesia, an inability to *spontaneously* generate bodily actions. A symptom often encountered in such patients is the "alien



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

hand sign" (Goldberg 1992). The hand contralateral to the lesion, typically the right hand, seems to take on a life of its own, without the control of the patient. In such patients the normal balance of MPS and LPS apparently shifts toward a dominance of the LPS. If an object is introduced into the intrapersonal space of a patient with the alien hand sign, the patient will grasp the object with such force that the fingers have to be prized off it. The relative role of the two systems in patients with MPS lesions is further shown in a study by Watson et al. (1986). They showed that such patients were maximally impaired in attempts to pantomime acts from verbal instruction. Less impairment was noted in attempts to imitate the neurologist's actions, and actual use of objects was most normal.

There are equivalent effects of MPS lesions for speech. The initial effect is often complete mutism – inability to spontaneously generate speech. However, subsequently, while spontaneous speech remains sparse, such patients typically show almost normal repetition ability. In these cases, Passingham (1987) has surmised that "it is Broca's area speaking" (p. 159). A similar pattern of results has been observed in patients with transcortical motor aphasia which typically involves interference with the pathway from the SMA to inferior premotor cortex (Freedman et al. 1984).

In contrast to these results of MPS lesions on speech are results of lesions of LPS, which tend to affect repetition more than spontaneous speech. In particular, this pattern is often observed in Conduction aphasics who tend to have damage in inferior parietal cortex affecting transmission of information from Wernicke's area to Broca's area. Thus the medial and lateral patients described here show a "double dissociation," a pattern much valued in neuropsychology because it provides evidence that there are two separable functional systems in the brain (Shallice 1988). Further evidence for this dichotomy comes from patients with "isolation of the speech area." These patients, who have lost most cortex except for lateral perisylvian cortex, have no spontaneous speech, but may repeat input obligatorily, without instruction (Geschwind et al. 1968).

### 6.4. The lateral system and speech learnability

Typical bodily actions are visually guided. While the motivationally based intention is generated in MPS, which may also help to provide the basic action skeleton, the action itself is normally accomplished, while taking into account target-related information available to vision by means of LPS. In contrast, spontaneously generated speech episodes are not sensorily guided to any important degree. However, as we have seen, the lateral system has an extremely good repetition capacity. Normal humans can repeat short stretches of speech with input-output latencies for particular sounds that are often shorter than typical simple auditory reaction times (approximately 140 msec; see Porter & Castellanos 1980). People have been puzzled as to why we possess this rather amazing capacity when, in the words of Stengel and Lodge-Patch (1955), repetition is an ability that lacks functional purpose.

A background for a better understanding of the repetition phenomenon comes from evidence from PET studies on the activation of ventral lateral frontal cortex (roughly Broca's area) in tasks that do *not* involve any overt speech;

> for example, the categorization of visually presented letters on the basis of their phonetic value (Sergent et al. 1992), a rhyming task on auditorily presented pairs of syllables (Zatorre et al. 1992), a sequential phoneme monitoring task on auditorily presented nonwords with serial processing (Demonet et al. 1992), the memorization of a sequence of visually presented consonants (Paulesu et al. 1993), a lexical decision task on visually presented letter strings (Price et al. 1993), and monitoring tasks for various language stimuli either auditorily or visually presented (Fiez et al. 1993). (Demonet et al. 1993, p. 44)

As Demonet et al. (1993) also note:

> The observed activation of this premotor area in artificial metalinguistic comprehension tasks suggests the involvement of sensorimotor transcoding processes that are also involved in other psychological phenomena such as motor theory of perception of speech (Liberman & Mattingly 1985), inner speech (Stuss & Benson 1986: Wise et al. 1991), the articulatory loop of working memory (Baddeley 1986), or motor strategies developed by infants during the period of language acquisition (Kuhl & Meltzoff 1982). (p. 44)

They note that the presence of this sensorimotor transcoding capacity is also suggested by "disorders of phonetic discrimination in Broca-type aphasic patients (Blumstein et al. 1977) as well as in subjects during electrical stimulations of the left inferior frontal region (Ojemann 1983)." (Demonet et al. 1993, p. 44)

The utility of this capacity and the probable reason for its origin becomes clearer when one notes that, while humans learn whichever one of the 6,000 or so languages they grow up with, monkeys have negligible vocal learning capacity (Jürgens 1995). The human repetition capacity is presumably associated with the now well established phonologic loop of working memory, which involves subvocalization as an aid in temporary storage of speech material (Baddeley 1986). Baddeley (1995) has recently speculated that this capability probably evolved in order for language to be learned. Thus, while in adults the primary role of the LPS for spontaneous speech is probably transmission of previously learned and now stored lexical information relevant to pronunciation from Wernicke's area to Broca's area, the primary role of LPS in infants is that it allows speech to be learned. It is somewhat ironic, in view of the special modular innate status often claimed for the human speech capacity, that from a perceptual-motor perspective the main change in vocal organization from other primates to humans may be evolution in the LPS of a capacity to *learn* speech. Furthermore, rather than having a unique form, the overall brain organization of motor output for speech seems to be no different than that for other bodily activity. Both are equally subject to the basic intrinsic-extrinsic functional dichotomy.

Lateral cortex presumably allows humans to not only say what they hear but do what they see, in general bodily terms. The presence of some ability of the SMA patients described by Watson et al. (1986) to imitate demonstrations of object use when they cannot pantomime such use is evidence of this; but there is also evidence that monkeys may possess some comparable ability. Pellegrino et al. (1992) have observed numerous instances in which single neurons in ventral lateral premotor cortex that had been shown to be active in various movement complexes performed by the animal also discharged when experimenters performed the same movements in front of the animal.

It seems likely that we have grossly underestimated the importance of our capacity for matching movements to input patterns, vested in the LPS, in our attempts to understand the evolution of cognition in general. Elsewhere, I


https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

have summarized an argument to this effect by Donald (1991), who believes that

> evolution of a generalized mimetic capacity in *Homo Erectus* was the first major step in the evolution of a hominid capacity beyond the great ape level, and was a necessary precursor to the evolution of language, which probably evolved in *Homo sapiens.* This hypothesis addresses the otherwise anomalous centrality in human culture of a wide range of behaviors including tribal ritual, dance, games, singing and music in general, all of which involve a capacity for the production of intentional representational displays but have virtually no analogs in living great apes. A wide variety of actions and modalities can be incorporated for the mimetic purpose: "Tones of voice, facial expressions, eye movements, manual signs and gestures, postural attitudes, patterned whole body movements of various sorts . . . " (p. 169). Donald makes the plausible argument that this mimetic capacity must have evolved before language, because language provides such a rich cognitive endowment that it would be hard to explain the necessity for mimesis once language had evolved. (MacNeilage 1994, pp. 186–87).

### *6.5. Speech input from posterior cortex*

Finally, with reference to speech, a word is in order about the input to the two proposed motor control subsystems. There is general agreement that perisylvian cortex in the temporoparietal region is involved in phonologic representations of at least the stem forms of many content words, especially nouns. In contrast, grammatical morphemes (function words and affixes) and perhaps aspects of verbs may be primarily controlled from frontal cortex, judging by the agrammatism that follows extensive lesions in lateral frontal cortex in classical Broca's aphasia. Most segmental serial ordering errors of speech in both normals and aphasics involve content word stems, not grammatical morphemes, and the F/C theory presented here is most relevant to content words.

Patients with lesions in temporoparietal cortex typically produce paraphasic speech – speech replete with segmental errors. Acoustic studies have shown that these errors for the most part are errors of choice of segments rather than errors in their motor control, the latter errors being more prominent in patients with ventral frontal lesions (MacNeilage 1982). From this, one can conclude that temporoparietal cortex is involved in phonologic encoding of lexical items – access to phonologic information about words and successful delivery of this information to the production control apparatus. It is hypothesized that this encoding phase involves two kinds of information, one kind for each of the motor control subsystems that have been discussed. Information regarding numbers of syllables in the word, suprasegmental information regarding stress placement, and perhaps information about vowels may be sent to the medial system for frame generation. Information about consonants and vowels may be sent to the lateral system for generation of content elements adjusted to their segmental context, as suggested earlier. According to this conception, the subsequent reintegration of the frame and content components must take place in lateral premotor cortex.

### *6.6. The role of prefrontal cortex*

The full story of the evolution of speech must include the history of selection pressures for communication in the context of overall hominid evolution. An important neurobiologic development in this regard is the enormous expansion of prefrontal cortex, a region involved in higher order organizational functions in general. Prefrontal cortex is heavily interconnected with the limbic system, leading MacLean (1982) to suggest that it affords "an increased capacity to relate internal and external experience" (p. 311). Deacon (1992) accords prefrontal cortex the primary role in the development in humans of a "low arousal" learnable communicative capacity independent of innate emotion-based vocalizations because of its "dominant status in the loop linking sensory analysis, emotional arousal and motor output" (p. 155). A specific functional linkage between dorsolateral prefrontal cortex and the SMA in humans was recently shown by Frith et al. (1991) in a study involving word generation as distinct from word repetition. Paus et al. (1993) also observed joint activation of prefrontal cortex with medial cortical sites during speech tasks. Studies cited earlier (e.g., Petersen et al. 1988) also suggest that prefrontal cortex may have played a dominant role in the evolution of grammar.

## 7. Some implications of the theory

### *7.1. Testability*

Is this theory testable? Predictions regarding levels of activity of the SMA and areas 6 and 44 in certain tasks, testable by means of brain imaging studies, can be made. One straightforward prediction is that mastication, sucking, and licking will involve more activity in ventral area 6 than in area 44, and more activity in area 44 than in the SMA. Another prediction involves the general claim that the SMA is specialized for frame generation and ventral premotor cortex for content generation. The prediction can be tested using artificial forms of speech that manipulate the relative role of the frame and content components: (1) "Reiterant speech," a condition in which segmental content demands are minimized but syllabic and phonatory demands are not, should produce relatively higher activity levels in SMA, and perhaps more activity in area 6 than in area 44. In this condition, the subject attempts to simulate words or utterances using only one syllable. For example, if the stimulus word is *concatenate*, the subject says "maMAmama," producing the same number of syllables as in the stimulus word with major stress on the second syllable. (2) Bite block speech or speech with the teeth clenched eliminates the demand on the mandible for syllabification but increases the demand on segmental production because every segment must be produced in an unusual way to compensate for the inability to adopt the usual jaw position for the sound. This condition should produce higher relative levels of activity in area 44 than in either area 6 or the SMA.

### *7.2. Comparison with other theories*

How does this theory compare with other current general conceptions of speech production that have implications for evolution? The concept of the syllable was found to be central in all the areas of subject matter considered in the formulation of this theory. With the exception of conceptions based primarily on evidence from segmental errors in speech (Shattuck-Hufnagel 1979; Dell 1986; Levelt 1989) this emphasis is not shared in other conceptions of speech



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

as a behavior. The syllable is given virtually no attention in two other current theories of the evolution of speech, the two-tubed vocal tract theory of Lieberman (1984) and the motor theory of speech perception (Liberman & Mattingly 1985). It is not mentioned in the most prominent conception of brain-language relations, the Wernicke-Geschwind scenario, reiterated by Damasio and Geschwind in 1984 (Also see Damasio & Damasio 1992). It scarcely figures in the most prominent current conception of the on-line control of speech – the articulatory phonology perspective of Browman and Goldstein (1986). It is incidental to what until recently has been the dominant conception of acquisition of speech production – the theory of Jakobson (1968). The F/C theory suggests that all of these approaches require drastic restructuring.

In other contexts, the syllable falls victim to a functional eclecticism that results in a lack of recognition that speech might be different from other functions because it has been subject to different selection pressures. For example, Norman and Rumelhart (1983) have constructed a model of typing control consistent with typical typing error patterns. The model is based on "the assumption that the motor control of a learned movement is represented by means of a motor schema, an organized unit of knowledge, differing from the form of knowledge widely studied in the literature on memory, language, and thought only in that it has as its output the control of body movements" (p. 55) (see also Rumelhart & Norman 1982). This eclectic approach to mental organization, unaffected by the possibility that different functions may be subject to different phylogenetic constraints, is relatively common in both cognitive science and neuroscience. The present contention is that no theory of either the organization of speech or its evolution that does not include the dual components metaphorically labelled frame and content in the present discussion is a viable one, whatever theories might be advanced to account for any other aspect of human function.

### 7.3. Other instances of the frame/content mode

In earlier writings my colleagues and I (MacNeilage 1987a; MacNeilage et al. 1984; 1985) had suggested that the F/C mode of phonology may have had a precursor in an F/C mode of bimanual coordination, in which the holding hand is the frame and the manipulating hand contributes content elements. I wish to retract this view because I was unable to conceive of an adaptation, induced by a specific selection pressure, that would have achieved the transfer of such a generalized organization capability from the manual to the vocal system. In this journal and elsewhere, my colleagues and I have suggested an alternative view of the hand-mouth relation whereby both the manual and the speech specialization arises from a left hemisphere specialization for whole body postural control already present in prosimians (MacNeilage et al. 1987; MacNeilage 1991b). It is possible that this role of the left hemisphere for whole body motor control may be fundamental to all vertebrates (MacNeilage 1997a). There seem to be other important F/C modes of complex behavior. Garrett (1988) has argued for an F/C mode of syntax on the basis of evidence from serial ordering errors involving morphemes and words. I regard this evidence as an extremely important clue as to the means of evolution of grammar, but would regard this mode as analogous to the F/C mode of organization of phonology rather than homologous. The F/C mode of organization can also be implicated in much hand-mouth interaction, such as that involved in one-handed and two-handed feeding (MacNeilage 1992). It appears that the F/C mode is an important means of evolution of complex action systems. Presumably this is because it makes it possible to produce a large number of output states with a small number of basic organizational configurations – one basic frame in the case of speech.

### 7.4. Innate subsegmental units

F/C theory provides no justification for the postulation of *innate* subsegmental units – either the feature of linguistic theory (Chomsky & Halle 1968) or its functional equivalent, the gesture (Liberman & Mattingly 1985). The concept of the feature, including the notion that it provides an innate basis for the "phonetic possibilities of man" (Chomsky & Halle 1968) arises from circular reasoning, as Ohala (e.g., Ohala 1978) has repeatedly pointed out; and I have discussed elsewhere the problem of providing independent evidence for the concept of gesture in on-line production and perception, let alone genetic structure (MacNeilage 1990). In addition to the inherent inability of essentialistic concepts, such as the concept of distinctive feature, to form part of a theory involving change, the concept is of no value in the present context because it lacks functional implications. For example, a distinctive feature such as "+high" (MacNeilage 1991a) refers to an end state of the tongue that can vary across a subclass of vowels. The characterization is considered to be abstract and only indirectly related to articulation (e.g., Anderson 1981). However, no coherent theory of the transforms from the putative single abstract representation to its various manifestations in vowels has ever been presented. The present *articulatory* connotations of the definitions of features ("high" refers to an articulator) introduced by Chomsky and Halle (1968) to replace the perceptually based features of Jakobson et al. (1951) were not chosen on the basis of any evidence regarding speech production, but only because articulatory terms had more straightforward connotations than perceptual terms. However, as a result of this decision, the discipline of phonology is now ill-equipped to describe, let alone explain, the many features of sound patterns of languages that apparently develop for perceptual reasons – for example the fact that nasals tend to assimilate to the place of articulation of adjacent stop consonants, but fricatives do not (Hura et al. 1994).

### 7.5. Input-output relationships

The F/C theory includes the suggestion that there have been major developments in the efficiency of input-output linkages in the evolution of speech. The motor theory of speech perception (Liberman & Mattingly 1985) also emphasizes the importance of the evolution of input-output relationships for speech. In this theory, the gesture is the fundamental unit of both input and output, with the abstract representation of the output unit serving as a basis for categorical perception of input. However, the motor theory calls for the opposite relation between phylogeny and ontogeny than the one suggested here. According to the motor theory (Liberman & Mattingly 1985), gestures originate as separate entities, and then, under pressures for rapid


https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

message transmission, become increasingly coarticulated with neighboring units to the point where only perceptual access to gestural invariance at some abstract production level makes them perceivable. If true, this would be a case in which ontogeny reverses phylogeny. As we have seen, the frame dominance state in babbling and early speech is characterized primarily by heavy coarticulation of successive articulatory positionings, and subsequent developments are in the direction of *reducing* coarticulation rather than increasing it (e.g., Nittrouer et al. 1989). Thus, rather than being the initial elements out of which speech was created, gestures, if they can be adequately defined, will probably be best regarded as later emergents, phylogenetically and ontogenetically. The F/C theory suggests instead that the syllable frame should be regarded as providing an initial common basis for interactions between perceptual, lexical, and motor subcomponents of the speech system in earlier hominids and modern infants.

### 7.6. Was the first language spoken or signed?

The question of whether spoken or sign language was the first language is considered in detail elsewhere (MacNeilage 1998a), with the following conclusions:

1. The current ubiquity of spoken language encourages a belief in its evolutionary priority. The reasons usually given for an historical switch from signed to spoken language – the lack of omnidirectionality of sign, the fact that it prevents other uses of the hands, and its lack of utility in the dark – seem insufficient to have caused a *total* shift from manual to vocal language.

2. The likelihood that there is a left hemisphere vocal communication specialization in frogs, birds, mice, gerbils, and monkeys, and the many instances of right handedness in groups of higher nonhuman primates (both reviewed in MacNeilage 1998b) casts doubt on the frequently encountered contention that tool construction and use in *Homo habilis* were crucial manual adaptations for language.

3. The repeatedly obtained finding that language lateralization is more closely related to foot preference – an index of postural asymmetry – than to handedness, which is an index of skill (Day & MacNeilage 1996; Elias & Bryden 1997; Maki 1990; Searleman 1980) casts further doubt on an early role of manual language.

4. Recent claims that there is a left hemisphere specialization for language independent of the modality (Poizner et al. 1987; Petitto & Marentette 1991) give the spurious impression that an historical shift from signed to spoken language could easily have occurred. These claims are found to be unjustified.

### 7.7. Coda

According to the F/C theory, the evolution of the control of the movements of speech from prespeech vocalizations involved preexisting phonatory capacities and a specific sequence of adaptations proceding from ingestive cyclicities, via visuofacial communicative cyclicities, to syllables, which ultimately became modulated in their internal content. The overall form of the theory is relatively straightforward, taking as it does a well accepted notion of the dual structure of speech organization (frames and content elements) and mapping it onto a relatively well accepted notion of the dual structure of primate cortical motor systems (medial and lateral) which were presumably modified for the purpose. It is hoped that this theory will provide an antidote, in addition to the one provided in this medium by Pinker and Bloom (1990), to the tendency to regard language as "an embarrassment for evolutionary theory" (Premack 1986, p. 133). My guess is that language will eventually prove to be amenable to current mainstream evolutionary theory. A neo-Darwinian approach to speech may prove to be the thin edge of the wedge for the understanding of language evolution.

ACKNOWLEDGMENTS

This paper was prepared with support from research grant No. HD-27733 from the Public Health Service. I thank Hugh Buckingham, Antonio Damasio, Barbara Davis, Randy Diehl, Rogers Elliot, Jean-Luc Nespoulous, Giacomo Rizzolatti, Marilyn Vihman, and Steven Wise for their comments on the manuscript. I also wish to thank Mario Liotti and Larry Parsons for generating Figure 3 by use of BrainMap software and Peter Fox for making the facilities of the Research Imaging Center, University of Texas Health Science Center, San Antonio, available for this purpose.

# Open Peer Commentary

*Commentary submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article. Integrative overviews and syntheses are especially encouraged.*

## Motor cortex fields and speech movements: Simple dual control is implausible

James H. Abbs[a] and Roxanne DePaul[b]

[a]*Department of Neurology, University of Wisconsin, Madison, WI 53705;* [b]*Department of Communicative Disorders, University of Wisconsin, Whitewater, WI 53012.* **jhabbs@facstaff.wisc.edu; depaul@waisman.wisc.edu**

**Abstract:** We applaud the spirit of MacNeilage's attempts to better explain the evolution and cortical control of speech by drawing on the vast literature in nonhuman primate neurobiology. However, he oversimplifies motor cortical fields and their known individual functions to such an extent that he undermines the value of his effort. In particular, MacNeilage has lumped together the functional characteristics across multiple mesial and lateral motor cortex fields, inadvertantly creating two hypothetical centers that simply may not exist.

We empathize with MacNeilage's attempt to synthesize such diverse opinions and data as those that surround the neurobiology of primate communication, infant babbling, and speech. Although many of his concepts are intriguing, we have difficulty with some of the critical details of cortical function.

First, we concur emphatically with MacNeilage's declaration that human speech has precursors in primate ingestive and communication functions. Many scientists have embraced this concept for some time (Abbs 1986; Abbs & Welt 1985; LeMay & Geschwind 1976; Lenneberg 1967; Peterson et al. 1978; Petersen et al. 1989; Premack 1976; Snowdon et al. 1982; Yeni-Komshian & Benson 1976). Snowdon et al. (1982) show particular insights on the tired "speech is special" argument; "The stress on the uniqueness of human language is an argument not much different from the creationist arguments of Darwin's time or the scientific



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

creationism" of today . . . [and is] inconsistent with modern evolutionary thought" (p. xv). Unfortunately, the intractable doctrine that oral communication is a wholly unprecedented human behavioral and biological phenomenon continues to hamper linguistic research and theory, as well as impeding more enlightened clinical management of speech disorders (Abbs & Rosenbek 1985).

On the other hand, we find MacNeilage's view of the "medial premotor system" (sect. 6.2) to be fatally over-simplified, especially given the large body of recent work in monkey and human cortex documenting multiple mesial neocortical motor centers. The traditional SMA (supplemental motor area) is now known to be comprised of 2 separate cortical fields, with 2–3 additional mesial motor fields in the traditional cingulate region (Deiber et al. 1991; Dettmers et al. 1995; Fink et al. 1996; He et al. 1995; Marsden et al. 1996; Picard & Strick 1997; Rizzolatti et al. 1996c; Tanji 1994; Wise et al. 1996). Each of these mesial cortical fields has different connections to other brain sites (Luppino et al. 1993; Matelli et al. 1991; Zilles et al. 1996), different response profiles to electrical stimulation (Fried et al. 1991; Luppino et al. 1991), and different activity patterns with simple and complex movements (Alexander & Crutcher 1990; Matsuzaka et al. 1992). Connectivity, activity, and cytoarchitectonics undoubtedly reflect differences in cortical field functions. It is important that not one of these more accurately defined mesial cortical fields appears capable of providing the sole support for the speech motor "frame" function postulated by MacNeilage.

In particular, the mesial cortical field that might contribute to MacNeilage's speech frame (a so-called PRE-SMA) does not appear to be in turn involved in motor control functions or motor sequencing per se. As noted, traditional SMA is actually two centers, a caudal "SSMA" (supplementary sensorimotor area) and a more rostral "PRE-SMA" (cf. Rizzolatti et al. 1996c). PRE-SMA (meaning not part of SMA) appears to be the site where vocalization is elicited electrically and where lesions lead to transient mutism or speech problems (Fried 1996; Fried et al. 1991; Kirzinger & Jürgens 1982). This PRE-SMA also appears to have projections to the medullary brain stem that might be adjacent to cranial motor nuclei (Keizer & Kuypers 1989). It is important to note that PRE-SMA also is called a "negative motor area (cf. Lim et al. 1994)," because its stimulation causes involuntary cessation of speech as well as other ongoing motor activities (Dinner et al. 1987; Fried 1996; Fried et al. 1991; Luders et al. 1988; Marsden et al. 1996). Finally, PRE-SMA activity is not correlated with voluntary movement details (Dettmers et al. 1995; Matsuzaka et al. 1992) and has cytoarchitectonic similarity and connections with prefrontal cortical fields (Lu et al. 1994; Luppino et al. 1993); PRE-SMA is therefore thought to be remote from motor programming (lacking connection to primary motor cortex) and more likely to be involved in overall "urge to move" or other global motivational functions, but not motor control per se.

In turn, the traditional SMA caudal region, now SSMA (or SMA proper), despite an accepted role in movement sequencing, a responsiveness to on-line sensory information, and involvement in complex movement programming, appears to have no speech-related functions. In lumping together the functional characteristics across SSMA and PRE-SMA, MacNeilage has inadvertently invoked a hypothetical mesial motor center that does not exist. A similar problem results in lumping SSMA (and PRE-SMA) together with cingular (more mesial) motor fields. That is, although monkey vocalizations are influenced by mesial lesions, these lesions often further involve third, fourth, or fifth cingulate motor areas (depending on the extent of the lesions), often including the SMA only incidentally (Jürgens 1979a; Sutton et al. 1974).

These considerations raise the possibility that mesial motor systems play little role in actual speech motor control and programming (cf. Damasio 1985). Although imaging studies indicate overall SMA activity during speech (Fox et al. 1996; Pardo & Fox 1993; Peterson et al. 1989), this presumably reflects PRE-SMA activation, related to global motivation to move, not to generation of speech movement sequences per se. A related interpretation of mesial cortical function is that is also involves extralinguistic vocalization, based on projections from pain or "emotion" centers (cf. Jürgens 1976; 1979b); hence its contribution to primitive vocalization in moneys and human infants. Indeed, based on known sensorimotor-based coordination of orofacial and laryngeal actions (Abbs & Gracco 1984; Gracco & Abbs 1989), lateral speech motor cortical sites (Abbs & Welt 198) may take over when vocalization becomes a linguistic vehicle for language; this is presumably when primitive vocalization/babbling behaviors drop out. Vocalization and even babbling-like sequences in response to electrical stimulation or lesions of mesial sites in humans may thus be a primitive residual.

Our position is therefore a refinement on the traditional view that cortical speech motor functions are primarily in the lateral precentral region. First, major projections from posterior parietal language areas (via the arcuate fasciculus) link to lateral precentral cortex, not to mesial fields (Abbs 1986; Galaburda & Pandya 1982; Gottchalk et al. 1983). Second, a significant body of research shows that the timing and coordination of speech movement sequences is guided moment-to-moment by sensory information, most probably at lateral sites (Abbs & Gracco 1984; Gracco & Abbs 1989). Third, ingestive functions that MacNeilage views as the evolutionary basis for speech motor control are also controlled in lateral fields.

MacNeilage also oversimplifies the lateral cortical site for ingestive cyclic behaviors ("masticatory center"), implying that it is coincident with Broca's area. There are two lateral orofacial centers outside the primary motor cortex involved in orofacial functions. Cytoarchitectonic analysis indicates Broca's homologue (Brodman's area 44) in Old World monkeys is not lateral to the motor face area, but in the more rostral depths of the arcuate sulcus (Galaburda & Pandya 1982). Muakkassa and Strick (1979) also found these two areas to be outside the motor cortex face area, Broca's area, and a second site lateral to the subcentral dimple. Hence, contrary to MacNeilage's suggestion, it is this second lateral site, not Broca's area, that is associated with rhythmic jaw and tongue movements in monkeys and humans (Foerster 1931; Lund & Lamarre 1974; Luschei & Goodwin 1975; Walker & Green 1938; Watson 1975).

Given this obvious dual control capability in the lateral cortex, MacNeilage's "frame" may arise from the homologue of this lateral masticatory center, with the so-called content being generated in the nearby Broca's area. Until the neurobiological viability of the frame/content abstraction is better established, however, such speculation is not warranted.

## A new puzzle for the evolution of speech?

Christian Abry, Louis-Jean Boë, Rafael Laboissière, and Jean-Luc Schwartz

*Institut de la Communication Parlée, BP 25, F-38040, Grenoble Cedex 9, France.* **abry@icp.inpg.fr**

**Abstract:** We agree with MacNeilage's claim that speech stems from a *volitional* vocalization pathway between the cingulate and the supplementary motor area (SMA). We add the *vocal self-monitoring* system as the first recruitment of the Broca-Wernicke circuit. SMA control for "frames" is supported by wrong consonant-vowel recurring utterance aphasia and an imaging study of quasi-reiterant speech. The role of Broca's area is questioned in the emergence of "content," because a primary motor mapping, embodying peripheral constraints, seems sufficient. Finally, we reject a uniquely peripheral account of speech emergence.

As nobody knows exactly what story has to be told about the emergence of speech, we prefer to consider the topic as a radical "puzzling puzzle." Because we agree with most of the pieces delineated by MacNeilage, we will take the opportunity of this commentary to pinpoint only the "missing links" that could reinforce Mac-



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

Neilage's puzzle. In accordance with the homology criterion, the global question is: What primate brain circuit is the *design* closest to that of the speech-language lateral premotor system (LPS), but also the medial premotor system (MPS) circuits? More specifically, are there building blocks, as in a volitional vocalization production control circuit or a vocal self-recognition monitoring system?

It seems to be generally agreed that nonhuman primate vocalizations are not under volitional cortical control. This is obvious when homologues of Broca's and Wernicke's areas (B-W) are lesioned, but not when the anterior cingulate gyrus (ACG) and its neocortical extension SMA are. We need to emphasise the *volitional* control of this pathway, clearly defined as *cingular,* or *instrumental* vocalization (Sutton et al. 1974). For us, this volitional aspect is the root (moreover a logically necessary condition) of the *vocal self-monitoring* system. Both systems have been explored by the same researchers, namely Jürgens (1992), Müller-Preuss (1983), and Ploog (1992). This animal model of the vocal self has inspired the human model for auditory hallucinations in schizophrenia (Frith 1992). One of its crucial features is the use of B-W circuitry to implement the *corollary discharge* hypothesis: this allows the recognition of one's own *expected* vocalizations. This is the first recruitment of the B-W system, the next being speech.

Having called attention to this missing link (volition and self monitoring, Abry 1997), we rejoin MacNeilage's suggestion that the "birth" of speech occurs with the sudden emergence of canonical babbling behavior at about 7 months of age under SMA control. We consider the 17 patients with SMA lesions reviewed by Jonas (1981) as "brothers" of Broca's "Tan-Tan." We add the clear set of eight global aphasic cases reported by Poeck et al. (1984; unfortunately without scanning the SMA region), with only one consonant-vowel recurring utterance and with labials or coronals just as in babbling. This reinforces the view that disintegration could somehow recapitulate ontogeny backward. In addition, current studies on dendritic growth have concluded that connections from Broca's area to the primary orofacial motor cortex develop at 12 to 15 months of age, thus rubbing out a Broca area as the basis for babbling emergence. Even in adults, Broca's area is clearly not needed to control respiration, phonation, open-close lip cyclicity, rhythm, and prosody, in "Buy Bobby a poppy," but SMA is! (Murphy et al. 1997).

Unlike pre-SMA or SMA proper, Broca's area's role remains mysterious for us (and for many others). This is not clarified by the beautiful and unique discovery of the perception/action system, called "mirror neurons," which gave rise to the latest proposal by Rizzolatti et al. (1996a) that neurons in a homologue of Broca's area could match the "observation" and the "execution" of visual lipsmack communicative cyclicities. For, provided that these affiliative lipsmacks could be volitionally controlled by the 3-day-old rhesus monkey (Redican 1975), *Broca's* area's control of this visual *and audio* "precursor," is clearly a misreading of the target article. We must still interpret recent data by Kim et al. (1997) that the pattern of activity in Broca's area is similar when two languages are acquired early and different when one is acquired later; there is no such difference found for Wernicke's area. We need to better understand how MacNeilage's mandibular *frame,* controlled by SMA, becomes "filled" by *content* units, that is, how the independence of the articulators carried by the jaw is achieved? Taking syndactylism as a test case, it seems that skillful independence is a matter of nonoverlapping primary cortical mapping in time. So what about the articulators? Is Broca's area necessary for such a mapping? When?

Finally, we do not think that control and neural structures fail to interact with the peripheral degrees of freedom and anatomical constraints to allow different forms of babbling to emerge: say [baba] for one baby, [dada] for another, in our simulations of different articulatory models. There is a need to correct a persistent error originating from Lieberman (1984) to the effect that speech has a *unique peripheral origin,* namely in the enlargement of the pharynx, which allows better motor control of the vocal tract and then a development of motor programs that lead ultimately to syntax. Recent simulations of a baby's vocal tract developed by Maeda and Boë (1997) clearly show that babies can produce an auditory space of the same magnitude in barks as adults, provided they can control [i] and [u] constrictions. Referring to MacNeilage (1994, p. 185): "The two-tubed vocal tract has increased our *articulatory* [our emphasis] possibilities. But the central question . . . is: How have we realized these articulatory possibilities in order to produce speech at rates as high as 15 phonemes per second . . . ?" with the proper emergence of babbling rate and independence of articulators.

## Cyclicity in speech derived from call repetition rather than from intrinsic cyclicity of ingestion

R. J. Andrew

*Sussex Centre for Neuroscience, University of Sussex, Brighton, BN1 9QG, England.* **bafe8@central.sussex.ac.uk**

**Abstract:** The jaw movements of speech are most probably derived from jaw movements associated with vocalisation. Cyclicity does not argue strongly for derivation from a cyclic pattern, because it arises readily in any system with feedback control. The appearance of regular repetition as a part of ritualisation of a display may have been important.

It is clearly true that human language is an evolved collection of abilities, subject throughout its evolution to selection pressures. It is impossible to see how it could have come into being instantaneously, and no biologist would believe that anything remotely like this ever happened. MacNeilage is to be applauded for saying this so clearly: evidently, it needs to be said regularly. He is also entirely right to examine the specific mechanisms that are basic to language and to consider in concrete terms how they might have evolved.

My criticisms are therefore intended to be constructive and are directed at a position that I regard as correct in its basic assumptions.

The argument neglects those jaw, lip, and tongue movements that are reflex components of vocalisation in mammals in general. It must be true that jaw movements are controlled by neural mechanisms that first evolved to allow biting: vertebrates had jaws long before they evolved the lungs they needed to vocalise. However, respiratory reflexes are at least as ancient as the tetrapods, and have been shaped by selection relating to vocalisation (amongst other things) for so long that it makes little sense not to consider their role in the calls of primates. In vigorous respiration, the entry of air is facilitated by mouth opening in inspiration; some degree of closure on expiration (together with disappearance of naris dilation) can be seen in some mammals (e.g., *Lemur,* Andrew 1963). It is likely that the original function of such partial closure was to prevent the lungs from emptying too rapidly, comparable to the glottal narrowing that occurs at the same time. However, a proper study of such movements of the jaws and face, both in respiration and in the calls that have evolved from respiratory reflexes, is badly needed, not least to answer the questions raised by MacNeilage.

It seems reasonable to start with the hypothesis that a single call delivered by an ancestral mammal commonly commenced with mouth opening, and ended with mouth narrowing or closure, accompanied by contraction of the *orbicularis oris.* Exactly such a sequence can be seen today in the *Lemur* and many other primates.

The next question concerns the origin of the cyclic repetition involved in the "frame/content" structure of human speech. MacNeilage argues that this requires the involvement of a neural control mechanism for jaw movements, which was already cyclic in



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

properties before the evolution of vowel/consonant alternation. However, cyclic repetition can readily be generated in any control mechanism that has feedback regulation of the end of the response. The appearance of intention tremor in Parkinsonism does not require the affected movement to be driven by a circuit that is designed to be cyclic or periodic in its functioning. Viewed this way, the rapid repetition of syllabic frames could represent repetition of the coordinations of a single call within one expiration. Such repetition is a common way of generating rapid call sequences in birds. It can be seen unambiguously in the domestic chick, where the same basic sound generating pattern, at the syringeal level, can either occupy a full expiration or be repeated through the course of a single expiration, with a momentary checking of thoracic emptying between each short call (Andrew 1964). It would be worth looking for repetitive jaw movements in bursts of rapid short calls in mammals.

The nature of the (hypothetical) feedback control is important. In love birds the forebrain system (*nu. basalis*), which controls jaw and tongue movements, receives independent and direct auditory information (Hall et al. 1994). Presumably, this is necessary to allow auditory input to be used to control the sound that is produced: Such movements are important in determining call properties. The interesting point, however, is that such direct input did actually evolve presumably through a stage in which its main function was only to control call characteristics and not to learn from fellows.

Primate calls, particularly those that have energy over a relatively continuous range of frequencies, and so (like baboon grunts) are well suited to reveal changes in vocal tract resonances, are extensively modulated by lip, jaw, and tongue movements (Andrew 1976). Once control of these movements by the properties of the sound being produced had evolved, at whatever point in human evolution, the ability to copy the sounds of others would have appeared, as well. The ability to use visual input to shape a grasp to the object to be grasped confers the ability to imitate hand postures in just the same way. It is misleading to discuss the general ability to mimic. Furthermore, it is premature to claim that sound mimicking is absent in primates other than humans (e.g., Masataki & Fujita 1989).

The movements that modulate baboon grunts include "lipsmacking," in which both jaw and tongue movements produce resonance changes comparable to those that cause the formant shifts that distinguish human vowels (Andrew 1976). The movements are regularly cyclic, much more so than the grooming movements from which they almost certainly derive. It is likely that this cyclicity is part of the evolution of a conspicuous display ("ritualisation"). If so, it is an example of the origin of cyclicity late in the evolution of a call, rather than an instance of its transfer from movements of ingestion. The existence of such modulation, which varies between cycles, makes it clear that it is not true that primate calls lack "independently variable internal subcomponents" (target article, sect. 2.1, para. 3).

## The missing link's missing link: Syllabic vocalizations at 3 months of age

Kathleen Bloom

*Department of Psychology, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada.* **kbloom@watserv1.uwaterloo.ca**

**Abstract:** If syllables are the link between nonhuman calls and human speech, as MacNeilage suggests, then that link is actually revealed in the "syllabic" sounds of the 3-month-old infant, well before the reduplicative babbling of the 8-month-old. Anatomical, acoustic, cognitive, and social perceptual evidence supports this earlier landmark.

From birth until the third month of life, the configuration of the human head, neck, and chest, as well as neuromotor and respiratory function relative to vocalization, resemble that of nonhuman primates. Neonatal (noncry) vocalizations are short in duration, simple in structure, and produced nasally with the mouth closed or slightly opened (Kent 1981; Netsell 1981). We have called these earliest vocalizations, "vocalics" (Bloom et al. 1987). With the growth of the head and neck by the third month, the vocal tract lengthens and bends, thereby uncoupling the epiglottis and velum, and infants begin to phonate orally. At the same time, the increasingly elliptical shape of the chest wall and increased neuromotor control of the intercostal muscles support longer phonations. Finally, the size of the tongue relative to the oral cavity decreases, and infants gain neuromotor control and independent movements of the tongue, jaw, and lips (Kent 1981; Netsell 1981). Together, these developments give rise to the physical capability to articulate during phonation. Segments have been identified in the intonational contours of sustained vowel phonation, and in the complexes of consonant-like and vowel-like sounds. We have called these vocalizations "syllabics" (Bloom et al. 1987). Segmentation in early vocalization has been described as "phrasing," and confirmed by perceptual and spectrographic recordings (Lynch et al. 1995).

Growth in cognitive attention also serves the development of articulation at 3 months of age. In advance of eye-hand-mouth coordination, young infants react to visual stimuli that attract their attention by moving their mouths and tongues (Jones 1996). The adult face and voice are the most powerful elicitors of the 3-month-old's attention and, together with the demands for caregiving and the infant's relative lack of mobility, yield prolonged and stereotypic bouts of face-to-face social interactions with adults. These episodes of eye-to-eye contact and adult "babytalk" to which the infant responds with mouth and tongue movements, may or not be accompanied by phonation. Meier et al. (1997) have described these "silent mandibular oscillations" in 8-month-old infants, but others (e.g., Netsell 1981) have observed them in 3-month-olds.

In short, as early as the third month of life, infants open and close their mouths and move their tongues while phonating with prolonged attention to the adult's face and voice. Thus, 3-month-old infants have both the ability and the opportunity to phonate in concert with jaw and tongue movements and thereby produce segmented vocalizations.

Adult reactions also suggest that syllabic sounds at 3 months are the first links between lipsmacks and speech. Three-month-old infants produced a higher percentage of syllabics when adults responded in "turn taking" patterns and with intonational word phrases ("Hi there, baby") as compared to nonverbal sounds (Bloom 1988). Thus, infant syllabic sounds were *elicited by verbal* communication. Second, mothers of 3-month-olds were more likely to vocalize in response to syllabics and to ignore vocalic sounds (Masataka & Bloom 1994). Thus, syllabic sounds *elicited* adult verbal communication. Third, adults attributed both greater social favourability and more *communicative intent* to infants when they produced syllabic sounds (Beaumont & Bloom 1993; Bloom & Lo 1990; Bloom et al. 1993).

Finally, both acoustic and visuofacial characteristics of syllabic sounds influenced adult perceptions. We created video tracks in which infants' faces appeared producing vocalizations with and without mouth movements. We dubbed the audio tracks of syllabic and vocalic sounds into the video tracks. Adults gave higher social favourability and communicative intent ratings to infants producing syllabic sounds and to infants whose mouths moved while vocalizing (Bloom 1993). The effects of mouth movements were attenuated with Japanese adults for whom visuofacial feedback of speech is culturally less relevant (Bloom & Masataka 1996).

To be sure, the syllabic sounds at 3 months do not include all of the phonetic contours that the infant will be capable of at 8 months, and 3-month-old syllabic vocalizing is not produced with the rate and rhythmicity of reduplicative babbling. Nevertheless, the convergence of anatomical, neuromotor, cognitive, and social



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

developments at 3 months of age is already sufficient to support production of syllabic sounds. Thus, the onset of syllabic sounds and their importance for the evolution of speech are evident at a much earlier age in human infant development than is commonly believed.

## Embodiment, muscle sense, and memory for speech

Hugh W. Buckingham

*Department of Linguistics, Louisiana State University, Baton Rouge, LA 70803.* **hbuck@salvador.speech.lsu.edu**

**Abstract:** MacNeilage's target article develops a theory for the evolution of human speech articulation along the lines of "slot-filler" structure. His content/frame schema commits him to the tenets of embodiment, muscle sense, and a memory for speech. My commentary ties these aspects together in their historical and current perspective.

There is a sense in which MacNeilage's frame/content evolutionary theory of the human speech capacity is timely and another sense in which it reflects clear historical continuity. A recent *BBS* target article on what memory is for (Glenberg 1997) made the case for considering embodiment as the basic ingredient for memory systems. In discussing phonology, Glenberg refers to George Lakoff's (1988, p. 40) notion that human phonological systems are in a very deep sense grounded in and ultimately constrained by the muscles, shapes, and control of articulation. In his *BBS* commentary Lakoff is considering the neural reality of connectionist accounts for human phonological systems, whereby "phonological processes, in large measure, would be characterized by conventionalized activation patterns controlling articulatory and acoustic processing." This, according to Lakoff, "would help to limit the general principles embodied in phonological patterns to those that are phonetically realistic." MacNeilage's point clearly implicates the effect of embodiment in the microgenesis of frames and content for speech production, and the embodiment is tightly linked to the underlying neural structures and their locations in the primate nervous system, which give rise to the oral/facial actions. Functional shift moves the teeth chatters, lip smacks, and so on over to the realm of articulation frame production and subsequent content proliferation within these frames. The memorial system for speech, therefore, is a typically embodied cognitive structure, arising as it did from "bodily interactions with the world." (Glenberg 1997, p. 1). Although articulatory actions are intransitive, they are actions nonetheless, and actions whose sensory/movement traces are set up over time, subsequently serving as the memory that ultimately serves to drive the production itself.

David Hartley (1705–1757), the 18th century British Association psychologist, and arguably the first physiological psychologist, developed the notion of "muscle sense" (Proposition 15 of *Observations on Man,* 1749). As they move about continually, bodily limbs lay down sensory traces through the internal kinesthetic system. These traces then serve as the representational patterns that link with the productive systems, eventually serving to call up the actions volitionally. This is of course at the heart of associationist psychology; the underlying notion of embodiment is implied throughout.

Approximately 100 years after Hartley's *Observations,* neurologists in the French tradition were grappling with the question of the faculty of articulate speech. Reports of articulatory speech disorders subsequent to stroke were on the rise and they required precise characterization of what it was that was lost. Because the patients who had severe articulatory disorders were still able to swallow, lick their lips, chew, and so on, and were in general free of lingual paralysis, systems specific to *speaking* had to be invoked. There was a certain consensus that the articulatory disorder compromised the "faculty of articulate speech." This was considered to be an "intellectual" faculty, which consisted of the "memory for the procedure one has to follow in order to articulate the words" (Broca 1861/1960, p. 54). This notion of embodied memory for speech did not actually originate with Broca; its seeds can be traced to Jean Baptiste Bouillaud (1825).

In sum, embodied memory is at the heart of many early attempts at characterizing the essential nature of the human speech system; it plays a major role in MacNeilage's frame/content theory, as well.

## The frame/content model and syntactic evolution

Andrew Carstairs-McCarthy

*Department of Linguistics, University of Canterbury, Christchurch, New Zealand.* **a.c-mcc@ling.canterbury.ac.nz**

**Abstract:** The frame/content theory suggests that chewing was tinkered into speaking. A simple extrapolation of this approach suggests that syllable structure may have been tinkered into syntax. That would explain the widely noted parallels between sentence structure and syllable structure, and also the otherwise mysterious pervasiveness of the grammatical distinction between sentences and noun phrases.

It is commonplace to emphasize that language must not be confused with speech, because speech is not the only medium through which language can be expressed. Although this statement is true, it is easy to be seduced by it into assuming something it does not entail, namely, that only cognitive or neurological developments are really important in the evolution of language. MacNeilage's focus on speech is an excellent antidote to any such cognitive bias, and his central argument, concerning the distinction between syllabic frames and their segmental content, is persuasive. I will suggest, however, that in its syllabic guise the frame/content distinction may have relevance beyond the control of articulation: it may also help to explain certain puzzling questions about why syntax is as it is.

In section 7.3 MacNeilage mentions Garrett's (1988) use of something like a frame/content distinction to explain syntactic production and indicates that he regards this as "an extremely important clue" to how grammar evolved. Yet he immediately qualifies this enthusiasm by saying that what goes on in syntax is analogous, not homologous, with what goes on in phonology. I suggest that MacNeilage is unnecessarily cautious here. Parallels between the structure of the syllable and that of the sentence have been remarked on by various researchers (Kaye et al. 1990; Levin 1985; Taylor 1996); moreover they had been noticed long ago by Plato in his *Theaetetus* (Ryle 1960). Broadly speaking, in languages with subject-verb-object word order, the relationships among sentence, subject noun phrase, predicate, verb, and nonverbal material in the predicate recall those among syllable, onset, rhyme, nucleus, and coda. More generally, in terms of the X-bar structure for phrases pioneered by Jackendoff (1977), heads, specifiers, and complements in phrases recall nuclei, onsets, and codas in syllables. Are these parallels just coincidental, or are they joint reflections of a grammatical architecture that is neutral between phonology and syntax? There is a third possibility: that basic sentence structure is an evolutionary derivative of syllable structure. The pervasiveness of the syntactic distinction between sentences and noun phrases (NPs) suggests reasons for taking this possibility seriously.

Is it inevitable that syntax should have evolved so as to distinguish one kind of structure, labeled "noun phrase," to which *John's arrival* and *The Niagara Falls* belong, from another kind, labelled "sentence," to which *John has arrived* and *It's raining* belong? It



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

is tempting to conclude that the answer must be yes. In philosophy and linguistics, sentences are seen as the prototypical syntactic mechanism for asserting propositions, that is, for saying things that may be true or false. NPs on the other hand, are the prototypical device for identifying the arguments of propositions, that is, for referring to things about which true or false statements may be made. But it is important to remember that, just as a sentence may fail to fit the world by being false, so an NP may fail to fit the world by having no reference (e.g., the NP "the present King of France" as uttered in the 20th century) (Russell 1905; Strawson 1950). So there seem to be two ways of failing to fit the world: (falsity and failure of reference) as well as two ways of fitting it (truth and reference). But why should there be precisely two ways rather than three or a dozen – or just one way, which we might call "(in)applicability"? The more painstakingly one seeks a basis for the truth-reference distinction, the harder it becomes to find one that is genuinely independent of the sentence/NP distinction that seemed at first sight to be motivated by it (Carstairs-McCarthy 1998; forthcoming). Could it be, then, that the truth/reference distinction is merely a byproduct of grammar?

If so (admittedly a big "if" at this stage), then the question of why the sentence/NP distinction is apparently universal in human languages becomes urgent. But MacNeilage's argument hints at an answer to it. When a need for a syntax arose (that is, when the vocal apparatus began to facilitate the fluent concatenation of individually meaningful calls), a neural mechanism for the control of a linguistically relevant activity was already at hand, namely, the mechanism for the syllabic structuring of speech. So it is hardly surprising that, syntax should even today display parallels with syllable structure, and that the kind of aphasia that tends to be accompanied by effortful articulation should be precisely Broca's aphasia, in which grammar is disrupted, rather than Wernicke's or jargon aphasia, where grammar is relatively intact. As MacNeilage says, "one needs to resist a tendency to regard mastication as too simple a candidate for tinkering into speech" (sect. 4.3). Similarly, one needs to resist a tendency to regard the neural control of syllable structure as too simple a candidate for tinkering into a blueprint for syntax.

## Distributed neural substrates and the evolution of speech production

Asif A. Ghazanfar and Donald B. Katz

*Department of Neurobiology, Duke University Medical Center, Durham, NC 27710.* **asifg@neuro.duke.edu; dkatz@neuro.duke.edu**
**www.neuro.duke.edu**

**Abstract:** There is evidence of reciprocal connectivity, similarity of oscillatory responses to stimulation of multiple motor and somatosensory cortices, whole system oscillation, and short-latency responses to behavioral perturbation. These suggest that frame/content may be instantiated by overlapping neural populations, and that the genesis of frame oscillations may be profitably thought of as an emergent property of a distributed neural system.

How separate can frame/content be in the brain? MacNeilage's extensive analysis locates them in distinct subdivisions of premotor cortex: frame specification and the internal generation of speech are produced in the medial premotor system (MPS), whereas content and stimulus responsivity are generated in the lateral premotor system (LPS; it is interesting that this is an area involved in the generation of ingestive cyclicities in primates). The suggestion that human speech production is evolutionarily derived from more basic orofacial behaviors is a welcome one. We elaborate on this idea here, in the process highlighting the importance of brain regions not discussed in the target article. Specifically, we suggest that the premotor cortices do not function in isolation from multiple motor and sensory cortical and subcortical areas, and that the speech frame may be best thought of as distributed in origin.

Although it is important that speech can be partitioned into frame and content (or into internally and externally driven) processes, it is not clear that the medial or lateral premotor cortical areas involved in this sensorimotor act are dedicated to only one of these modes. In primates, studies comparing supplementary motor (SM) with premotor (PM) cortical neuronal activity demonstrate that their functions overlap extensively when explicitly testing visually- versus internally-guided movement generation (Kurata & Wise 1988; Mushiake et al. 1991). Mushiake et al. (1991) concluded that "neither SMA nor PM is exclusively related to any particular motor behavior" (p. 716). Despite distinctions made on the basis of hodology, neural activity, and lesion evidence, the specification of a singular functional role for each of these cortical areas remains somewhat controversial. [See also Goldberg: "Supplementary Motor Area Structure and Function" *BBS* 8(4) 1985.]

A variety of anatomical, physiological, and behavioral evidence further suggests that at least the frame process may be distributed across a broader neural circuit than premotor cortex. Primate medial and lateral premotor cortices have extensive overlapping and reciprocal connections with other cortical areas involved in orofacial movements (Luppino et al. 1993; Matelli et al. 1986). This interconnectivity is reflected in the fact that rhythmic jaw movements can be elicited by microstimulation of primary motor (MI) or somatosensory (SI) cortices (Huang et al. 1989). The firing rates of many MI and SI cortical neurons are modulated during mastication and swallowing, as well as during the voluntary control of tongue movements (Martin et al. 1995). It is conceivable that such activity, together with premotor activity, could underlie the articulatory gestures of some macaque vocalizations (Hauser et al. 1993).

A speech act involves coordination of the entire motor neuraxis, and it may ultimately be worthwhile to consider frame specification in terms of a distributed system mechanism. As noted by MacNeilage, there is ample evidence that spinal cord and brainstem circuits can produce functional motor rhythms on their own (Pearson 1987). The mandibular system itself probably needs little prompting to settle into rhythmic behavior, and in human infants it begins to do this before the proposed cortical hardware, the medial premotor system, is mature (Blinkov & Glezer 1968). Furthermore, motor oscillations in general (and speech rhythms in particular) seem to be too flexible, robust, and quick to adjust to perturbation to be explained in terms of purely cortical control. Compensation for perturbations of mandible movement during speech can occur throughout the speech apparatus in as little as 20 to 30 msec (Kelso et al. 1984), a fact that implies both interaction between frame and content production and, at least partially, subcortical control of speech.

Sensorimotor neural rhythms have been detected in humans using scalp electroencephalography (EEG) and magnetoencephalography (MEG), and in animals using a variety of techniques. Chronic and simultaneous recordings of neuronal ensembles have revealed synchronous firing of neurons distributed across rat somatosensory and motor cortices, somatosensory thalamus, and the trigeminal complex (Nicolelis et al. 1995; 1997). Such synchronous activity has been found to predict small amplitude whisker movements (Nicolelis et al. 1995). Although these oscillations were first detectable in cortex, the phase relationships of firing among the different levels of the system suggested that the system assembled itself into an oscillating unit, without a localized driving source. Such behavior is typical of physical systems operating with continuous feedback and has been observed in biological neural networks (Marder & Calabrese 1996). It could be argued that the CPG (central pattern generator) for movement – the frame of the frame/content model – is as ubiquitous as it seems because it is a general property of distributed neuromuscular systems, as opposed to being reduplicated in multiple locations (Thelen 1991). [See also Selverston: "Are Central Pattern Generators Understandable?" *BBS* 3(4) 1980.]



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

MacNeilage should be commended for his formulation of a theory of speech production that makes good sense from an evolutionary perspective. It is one of the few hypotheses that can be tested explicitly using comparative neurobiological data and functional brain imaging. However, the physiological and behavioral evidence we have cited points to neural substrates beyond the premotor cortices that could be involved in speech production. It is unlikely that a given locus is solely responsible for one component of speech output.

## Premotor systems, language-related neurodynamics, and cetacean communication

Gary Goldberg and Roberta Brooks

*Drucker Brain Injury Center, Moss Rehab Hospital, Albert Einstein Healthcare Foundation, Philadelphia PA 19141-3099.* **goldberg@vm.temple.edu; rbrooks@aehn2.einstein.edu www.einstein.edu/phl/**

**Abstract:** The frame/content theory of speech production is restricted to output mechanisms in the target article; we suggest that these ideas might best be viewed in the context of language production proceeding as a coordinated dynamical whole. The role of the medial premotor system in generating frames matches the important role it may play in the internally dependent timing of motor acts. The proposed coevolution of cortical architectonics and language production mechanisms suggests a significant divergence between primate and cetacean species corresponding to major differences in areal differentiation trends in cerebral cortex.

MacNeilage has synthesized some very interesting ideas about brain evolution, cortical structure, and language development on both phylogenetic and ontogenetic scales. We would like to comment briefly about the importance of linking the dynamics of speech action and the rhythmicity inherent in speaking to the dynamics of the internal linguistic processes that culminate in overt language behaviors. The rhythmic structure of the overt behavior must be smoothly linked to those of the closely coordinated processes of semantic search, phonologic encoding, and syntactic sequence generation that are the focus of much of language-related cognitive science and that MacNeilage chooses to leave out of the frame so as to concentrate on the motor control output issues. It is difficult to isolate output from underlying processes; indeed, the study of output generation can lead to important inferences about how the underlying neurodynamic processes may be constrained. In considering the differentiation of premotor systems that are intrinsically generative from those that are extrinsically responsive, it becomes apparent that the source of the timing that allows anticipatory or projectional behavior (e.g., propositional language) must be internal, whereas the temporal dynamics of responsive behavior can occur through links to external rhythmicity.

There is accumulating evidence for an internal clocking mechanism that operates at ultradian frequencies and plays an important role in governing the timing of self-generated action. This appears to be closely associated with structures in the medial premotor system (MPS), including the supplementary motor area (Lang et al. 1990). Patients with Parkinsonism, for example, who have evidence of impaired function of the MPS, have been found to have abnormal internal clock functions that can be improved with dopaminergic medication (Pastor et al. 1992). This may be related to a decreased rate of dopamine-dependent activation of the MPS-related regions of cerebral cortex, including the supplementary motor area, prior to the emergence of a self-generated motor act (Dick et al. 1987; Jahanshahi et al. 1995). Language output in Parkinsonism can be facilitated through the provision of an external source of rhythmic pacing. That the MPS generates the frames suggests that its primary involvement is in controlling the temporal envelope for the production process, which is consistent with the internal "clocking" idea. Thus, the rate of frame production is a critical "clocking" variable that must serve to synchronize the entire production process in real time. The chronogenetic tradition in aphasiology is based on the general idea that speech errors are more likely when this synchronization fails (Kolk & van Grunsven 1985). MacNeilage considers the issue of frame rate and rhythmicity in section 5.4 (our intuition from supermarket shopping, though, has failed us so far). We propose that this production rhythmicity has deeper importance based on the need to entrain and synchronize *all* of the underlying spatially distributed processes so that speech production can proceed as a coordinated dynamic whole. One of the time-critical parts of the production process is semantic search of the lexicon; when impaired, this results in semantic paraphasias (e.g., substituting "bench" for "chair"), a problem that may be viewed as a premature surfacing of the search process. This can occur as the result of a mismatch between frame production and semantic search dynamics. Similarly, if frame rate exceeds the ability to fill in a frame with the correct phonologic content, then phonemic paraphasias can occur again as a result of a mismatch between frame production and phonological search dynamics. When patients are able to monitor and correct errors, one important way they may compensate to reduce error rates is by slowing down the rate of speaking (Kolk 1995). When patients lack an awareness of their own errors, the production rate may actually drift upward with an associated increase in errors.

Schwartz et al. (1994) and Martin et al. (1994) have addressed speech production errors in normal and aphasic individuals in the context of a nonlinear dynamical system: a spreading activation model (Dell 1986), which places specific temporal constraints on convergent inputs from sequentially activated semantic, word, and phonological nodes that vary with the networks' signalling efficiency. Signalling efficiency is viewed as the relative rate at which activation can be accurately delivered to the intended targeted nodes relative to nontargeted nodes. Efficiency is measured as the error rate as a function of speech rate. Patterns of errors in jargon aphasics can be duplicated in normals by making them speak more quickly. Improved efficiency in normals can be obtained through a strengthening of connections along an internodal pathway to the targets achieved through task practice. Loss of efficiency can result from damage to the network that weakens internodal connections. Furthermore, an inability to sustain the activation delivered to the target through to the point of "read-out" can lead to a different type of error profile (Martin et al. 1994). Thus, the matching of speech rate to signaling efficiency is important in minimizing error rate and suggests the importance of the selection of frame rates in ensuring correspondence between speaking-rate- and production-related neurodynamics.

With regard to the evolution of species-specific vocalization and communication, the close relationship between the emergence of human language and the elaboration of medial and lateral gradations in the phylogenetic development of mammalian cortical architectonics suggests possible insights into one of nature's most interesting experiments: the return of mammalian species from terrestrial to aquatic life in the cetacean forms. These animals have evolved a distinctive capacity for vocalization and, presumably, communication, without the parallel evolution of distal limb dexterity that also appears to have important ontogenetic linkages (e.g., Locke et al. 1995). What makes this so interesting is the comparative anatomy of the cetacean brain: the cerebrum of the cetacean brain is massively corticalized to a significantly greater extent than in humans, yet the architectonics show significantly less areal differentiation and certainly nothing similar to the distinctive primary motor cortical structure found in the human precentral gyrus (Morgane et al. 1985). What could this imply about the nature of the structure and function of vocalization and communication in these advanced aquatic creatures, or about their intelligence relative to our own? It may well be difficult to identify a clear frame/content structure in cetacean vocalization or to determine the function of cetacean vocalization if we attempt to do



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

so while constrained by the framework in which we have come to understand human vocalization and language function.

## A syllable-centric framework for the evolution of spoken language

Steven Greenberg

*International Computer Science Institute, Berkeley, CA 94704.*
**steveng@icsi.berkeley.edu www.isci.berkeley.edu/steveng**

**Abstract:** The cyclic nature of speech production, as manifested in the syllabic organization of spoken language, is likely to reflect general properties of sensori-motor integration rather than merely a phylogenetic progression from mastication, teeth chattering, and lipsmacks. The temporal properties of spontaneous speech reflect the entropy of its underlying constituents and are optimized for rapid transmission and decoding of linguistic information conveyed by a complex constellation of acoustic and visual cues, suggesting that the dawn of human language may have occurred when the articulatory cycle was efficiently yoked to the temporal dynamics of sensory coding and rapid retrieval from referential memory.

The syllable is an important representational unit that has largely been neglected in models of speech perception/production and spoken-language understanding. In many ways the syllable serves as the interface between sound and meaning (Greenberg 1996; 1997); it is refreshing for an evolutionary perspective, such as MacNeilage's, to afford a central role to this important unit of linguistic organization. This commentary focuses on the evolutionary origins of the syllabic cycle in speech production and the importance of "information" and "time" for sculpting the contours of this modulatory activity.

MacNeilage suggests that the origins of speech production may be linked to mastication, which bears a motor similarity to the open and closing phases of the articulatory cycle associated with syllabic elements of spoken language. Although there may indeed be some evolutionary relation between this nonlinguistic, motor behavior and speech, an alternative perspective, based on the temporal properties of sensorimotor function and integration, provides a potentially more comprehensive and explanatory framework with which to investigate the evolutionary conditions under which spoken language arose.

The time interval corresponding to the *average* length of a syllable – 165 to 200 msec (Arai & Greenberg 1997; Greenberg et al. 1996) – is ubiquitous with respect to neurological function, corresponding to the time constant for energy integration in both audition (e.g., Eddins & Green 1995) and vision (e.g., Regan & Tyler 1971), as well as to the minimum response time for motor activity (e.g., Meijers & Eijkman 1974). This interval also corresponds to the time required by many regions of the cortex to classify and evaluate sensory events (e.g., Rohrbaugh et al. 1974) and to retrieve pattern-relevant information from memory (John 1967). The temporal properties of the articulatory cycle are likely to reflect this general sensorimotor and information-retrieval integration time constant.

The syllabic structure of spoken language is more complex and heterogeneous than MacNeilage's characterization implies. A syllable in English can assume 1 of approximately 15 different segmental variations with respect to consonant-vowel (CV) composition and order (Greenberg 1997). Although the CV form favored by MacNeilage is the most common variant (34% of the phonological forms, 47.2% of the phonetically realized instances), other syllabic patterns, such as CVC (31.6%, 22.1%), VC (11.7%, 4.8%), and V (6.3%, 11.2%), occur quite frequently (Greenberg 1997). Together, these 4 syllabic forms comprise 83.6% of the phonologically defined (and 85.3% of the phonetically realized) syllables in a corpus of spontaneous (American English) discourse (Switchboard, cf. Godfrey et al. 1992). The remaining 16.4% (14.7%) of the syllables reflect more "complex" forms containing consonant clusters at either onset, coda, or both. Although these complex syllables comprise less than a sixth of the corpus, their importance should not be underestimated. Most of these forms are associated with low-frequency content nouns (such as "strength" [CCCVCC] or "flasks" [CCVCCC]), which provide much of the informational detail characteristic of spoken language. This heterogeneity in phonetic composition is reflected in the variability of syllabic durations. Although the mean duration of a syllable is 165–200 msec, the standard deviation of this distribution is high (about 100 msec, for both English and Japanese, indicating that 85% of the syllables vary between 100 and 300 msec in length [Arai & Greenberg 1997; Greenberg et al. 1996]), reflecting the heterogeneous segmental composition of the syllabic elements. This variability in syllabic duration is significant for understanding the neurological bases of information coding in spoken language. Commonly occurring words (over 80% of which are monosyllabic), largely predictable from context (e.g., "function" words), tend to be pronounced in a "reduced" fashion closer to the canonical CV structure rather than low-frequency, highly informative "content" words. Deviation from this canonical pattern appears to be one means of linguistically marking elements invested with unusually high entropy.

Thus, the information associated with any specific linguistic element is likely to be reflected in its duration; hence the temporal properties of speech production provide a potential window onto the neurological mechanisms mediating the lower and higher levels of spoken language. The distribution of syllabic durations (in both English and Japanese) matches the low-frequency modulation spectrum (defined as the magnitude of energy in the speech signal low-pass filtered below 20 Hz, cf. Greenberg & Kingsbury 1997), with a peak at about 5 Hz (reflecting the mean syllabic duration of 200 msec) and substantial energy distributed between 3 and 10 Hz (Arai & Greenberg 1997; Greenberg et al. 1996). This modulation spectrum corresponds closely with the temporal transfer function of neurons in the AI region of primary auditory cortex (Schreiner & Urbas 1988) and the pattern of vocal movements during continuous speech (Boubana & Maeda 1998; Smith et al. 1993).

Together, these data suggest that the temporal properties of spoken language may not merely reflect constraints imposed by the inertial characteristics of a biomechanical system descended from a phylogenetically more basic (masticatory) function. They also represent the integration of the articulatory apparatus into an intricately woven web of sensorimotor function optimized for rapid retrieval of stored information that also underlies the brain's capability to construct a stable representation of the external world under the wide range of environmental conditions typical of the real world.

## Content without a frame? The role of vocabulary biases in speech errors

Trevor A. Harley

*Department of Psychology, University of Dundee, Dundee, DD1 4HN, Scotland.* **t.a.harley@dundee.ac.uk**
**www.dundee.ac.uk/psychology/staff.htm***#harley*

**Abstract:** Constraints on the types of speech errors observed can be accounted for by a frame/content distinction, but connectionist modeling shows that they do not require this distinction. The constraints may arise instead from the statistical properties of our language, in particular, the sequential biases observed in the vocabulary. Nevertheless, there might still be a role for the frame/content distinction in syntactic planning.

Despite many recent advances in our understanding of speech production, much is still uncertain about how a nonlinearized semantic plan is translated into a serially ordered phonetic message. A distinction between frames and content is a convenient way of accounting for serial order effects in language production, and


https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

MacNeilage provides a great service in focusing on this distinction. In doing so, however, he over-stresses the evidence for the distinction at the segmental level, while glossing over the extension of the theory to the syntactic level.

First, connectionist modelling has shown that the phenomena traditionally interpreted as necessitating a frame-content distinction have alternative interpretations. Early connectionist and connectionist-type models that were based on spreading and interactive activation certainly relied on the computational convenience of the slot-filler formalism to circumvent the problems of serial ordering (e.g., Dell 1986; Harley 1993; Harley & MacAndrew 1995). In contrast, new models indicate how this reliance can be avoided. For example, Vousden et al. (1997) use a series of oscillators to provide the timing mechanism for phonological encoding in speech production. This model demonstrates the observed range of speech errors – including segment movement errors – and accounts for the constraints on segmental speech errors discussed by MacNeilage without making use of an explicit frame-content distinction. In particular, our model demonstrates the syllable position effect, whereby when segments exchange in speech errors, they come from similar positions within a syllable.

In a similar vein, Dell et al. (1993) used connectionist modelling to show how some of the constraints on segmental speech errors might arise as a consequence of the statistical properties of the language rather than as a consequence of constraints attributable to the frame-content distinction. In particular, Dell et al. examined the behaviour of a recurrent network (Elman 1990) that learned to produce sequences of phonological features. The consonant-vowel category effect (whereby consonants substitute only for other consonants in speech errors, never for vowels, and vice versa), the syllabic constituent effect (whereby vowel-consonant sequences are more likely to be replaced than consonant-vowel sequences in speech errors), and the initial consonant effect (where initial consonants are more likely to be involved in speech errors than other segments) are readily explained by distinguishing between phonological frames and segmental content. Dell et al.'s model, however, could account for these effects without an explicit division between frame and content, because of the sequential bias inherent in the language vocabulary. (The model could only produce nonmovement errors.)

Of course, this begs the question of how this sequential bias arose, and MacNeilage's proposal might well be germane here. Nevertheless, just because a bias arose through a particular mechanism, it does not follow that the organism still uses that mechanism, or that a derivation of it, such as frames and fillers, is used in processing. Modelling shows that there are alternatives.

Second, speech error evidence motivates a distinction between lexical items and a syntactic plan that is best conceptualised as a frame (e.g., Garrett 1975). In this approach, open-class items are inserted into the grammatical framework for the planned utterance; closed-class items are immanent in this frame. It is an obvious extension of MacNeilage's proposal to cover syntactic planning in the same way as phonetic encoding. Indeed, in many ways syntactic planning is a more appealing candidate for the frame-content approach then segmental planning, because alternative explanations of the linearization process are far less clear. Whether this approach will eventually yield to a similar sort of connectionist resolution as that of segmental planning, and whether it will yield to the biological approach proposed by MacNeilage, remain to be seen.

## Speech evolved from vocalization, not mastication

Uwe Jürgens

*German Primate Center, Kellnerweg 4, 37077 Göttingen, Germany.*
**ujuerge@gwdg.de**

**Abstract:** The segmentation of phonation by articulation is a characteristic feature of speech that distinguishes it from most nonhuman vocalizations. However, apart from the trivial fact that speech uses some of the same muscles and, hence the same motoneurons and motorcortical areas used in chewing, there is no convincing evidence that syllable segmentation relies on the same pattern generator as mastication. Evidence for a differential cortical representation of syllable segmentation ("frame") and syllable "content" is also meager.

MacNeilage's frame/content theory has merit in that it draws attention to articulation as the most important discriminative feature distinguishing human from nonhuman primate vocal communication – as far as it concerns the motor side. MacNeilage concludes rightly that speech evolution depends on articulatory capacity. This commentator cannot follow MacNeilage's attempts to explain articulation as a derivative of mastication, however, either behaviorally or neurobiologically. Behaviorally, mastication is characterized by a rhythmic jaw movement with a frequency of about 1.5 Hz (in humans). Syllable production occurs with a frequency of about 5 Hz, and the syllable segmenting consonants are not always accompanied by jaw movements: for example, velar consonants between identical vowels, such as [eke], labial nasals [m] between vowels such as [u] and [o], certain dental stop-vowel combinations, such as [idi], and so on. MacNeilage even goes as far as assuming "that speech makes use of the same brainstem pattern generator that ingestive cyclicities do." (sect. 4.3). The fact that masticatory and articulatory rhythm generators have very different cyclicities, and that articulation, in contrast to mastication, does *not* consist of rhythmic jaw movements with a rather constant frequency, clearly weakens MacNeilage's theory.

Neurobiological evidence also seems to contradict MacNeilage's hypothesis. The pattern generator for mastication is located in the Nucl. reticularis gigantocellularis at the caudal pontine level (Nakamura & Katakura 1995). From here, masticatory motoneurons in the trigeminal motor nucleus are activated indirectly via the medullary parvocellular reticular formation. In the squirrel monkey there is a vocalization type, called cackling, that comes close to a multi-syllabic human utterance in that it consists of repetitively alternating harmonic and nonharmonic elements in a 10 Hz rhythm with involvement of supralaryngeal muscle activity in the rhythm of the repetition rate (Jürgens 1979; Kirzinger & Jürgens 1994). If during the production of cackling calls the nucl. reticularis gigantocellularis is stimulated electrically via an implanted electrode, there is no change in the rhythm of cackling (Dressnandt & Jürgens 1992). If, on the other hand, the parvocellular reticular formation is stimulated during cackling, a severe deterioration of the acoustic structure (including rhythm) occurs (Dressnandt & Jürgens 1992). This suggests that masticatory and vocal rhythmic pattern generators reside in different brain areas.

Problems also arise at the cortical level. MacNeilage assumes that the "frame" and "content" of syllabic utterances are controlled by different cortical regions, namely, the lateral inferoposterior frontal cortex (corresponding to the face area of the motor and premotor cortex) in the case of "content," and the dorsomedial frontal cortex (i.e., supplementary motor area and anterior cingulate gyrus) in the case of "frame." Because the "frame" is said to provide the syllabic segmentation of the utterance, it should be expected that lesions in the dorsomedial frontal cortex affect the rhythm of syllable production. This is not the case, however. The characteristic syndrome after lesions in the dorsomedial frontal cortex is transcortical motor aphasia, a state in which there is a low motivation to initiate speech with retained ability to repeat sentences spoken to the patient (Rubens 1975). Disturbances in



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

speech rhythm have been reported, however, after lesions in the face area of the motor and premotor cortex (corresponding to MacNeilage's "lateral frontal system") as well as after lesions of the cerebellum, a structure projecting to the facial motor cortex via the ventrolateral thalamus (Darley et al. 1975). This suggests that syllabic "content" and syllabic "frame" are both controlled by the lateral posteroinferior frontal cortex. Such an interpretation would also be in harmony with the somatotopical representation in the cortex. According to MacNeilage, whereas the "frame" is determined primarily by movements of the jaw, the "content" is predominantly shaped by the lips and tongue. Because all three organs are represented in the lateral posteroinferior frontal cortex (Foerster 1936b), and both syllable segmentation ("frame") and syllable shaping ("content") involve jaw as well as lip and tongue movements, it appears very unlikely that "frame" and "content," as MacNeilage uses these terms, are controlled by different brain regions. The argument that electrical stimulation of the supplementary motor area produces perseverative syllable repetition – an argument taken by MacNeilage to underline the role of this region in syllable segmentation – is convincing only on the first view. Electrical stimulation is a very artificial form of interaction with a specific brain region in which activation and interference with normal neuronal activity take place. In other words, electrical stimulation does not produce the normal behavior controlled by the stimulated site. Because the supplementary motor area is generally assumed to be involved in the initiation of complex motor patterns (Eccles 1982), the repetitive utterance of the same syllable might be interpreted as the result of simultaneously activating a speech-initiation mechanism and interfering with the normal speech flow. In other words, it is conceivable that the rhythmic syllable repetition during supplementary motor area stimulation is an artifact rather than the normal function of this area.

It may be added that one of the experiments MacNeilage proposes to test his theory has already been done in the manual rather than the oral domain. According to MacNeilage's frame/content theory, reiterant speech (i.e., uttering the same syllable repetitively) should activate the supplementary motor area more heavily than the lateral posterior frontal cortex. In an experiment by Roland et al. (1980) comparing (1) regional cerebral blood flow (rCBF) during repetitive flexion of the index finger against a spring-loaded movable cylinder with (2) rCBF during a complex sequence of learned finger movements, it turned out that during the simple repetitive movement, there was an activation in the hand motor area, that is, lateral posterior frontal cortex, but no activation in the supplementary motor area. Only during the complex finger movement sequence were the lateral frontal cortex and supplementary motor area activated. If (1) simple repetitive finger movements are comparable to repetitive utterance of the same syllable and (2) complex finger movement sequences are comparable to variable syllable sequences, Roland et al.'s finding would be counter to MacNeilage's prediction.

## Is the syllable frame stored?

Willem J. M. Levelt and Niels O. Schiller

*Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands.* **pim@mpi.nl; schiller@mpi.nl**

**Abstract:** This commentary discusses whether abstract metrical frames are stored. For stress-assigning languages (e.g., Dutch and English), which have a dominant stress pattern, metrical frames are stored only for words that deviate from the default stress pattern. The majority of the words in these languages are produced without retrieving any independent syllabic or metrical frame.

The starting point for MacNeilage's frame/content approach to the evolution of speech production is the evidence from speech errors. Segmental errors tend to preserve syllabic constraints: onsets tend to exchange with onsets (*m*ell *w*ade), nuclei with nuclei (b*u*d b*e*gs), and codas with codas (go*d* to see*n*). The pioneers of speech error research therefore suggested that during phonological encoding, speakers do not retrieve word forms as unanalyzed wholes, but rather as a sequence of segments to be inserted into a metrical frame (Fromkin 1971; Garrett 1975; Meringer & Mayer 1895; Shattuck-Hufnagel 1979). Much ink has been spilled on the nature and the origin of such a metrical frame. Although syllabic frames have been quite prevalent in the speech error literature, they have always been part of larger word-spanning frames. In most theories, it is these metrical frames for words that are independently retrieved from the mental lexicon. That idea was captured in the conclusion cited in the target article (sect. 3.2) from Levelt (1989): "A word's representation in memory consists of components that are relatively accessible and there can be metrical information about the number and accents of syllables without these syllables being available." Theories differ in how much metrical detail they specify for the stored word frame. It makes little sense, for example, to store all of a word's syllable frames if they are all of the general kind Onset-Nucleus-Coda (with potential null-segments as fillers); knowing how many there are in the word should be enough. If syllable-internal information is to be specified, it should contain more detailed CV-information, as is the case in one version of Dell's theory (Dell 1988).

Still, more recent work in phonological encoding, in particular reaction time research, suggests that, at least for stress-assigning languages such as Dutch and English, the metrical word frames are quite meager. They probably do not specify more than the word's number of syllables and the word's main stress position (Levelt 1992; Roelofs & Meyer, in press). That makes sense, because retrieving detailed syllable frames from the lexicon would be rather counterproductive. A lexical item's ultimate syllabification in connected speech is strongly context-dependent; for example, syllables often straddle lexical boundaries (cf. Levelt 1992; Schiller et al. 1996). More surprisingly, recent experimental evidence obtained by Meyer and Roelofs (cf. Levelt et al., in press) shows that stored word frames play a role only for words that have nondefault stress (in both Dutch and English a word with default stress is one with stress on the first syllable having a full vowel). In Dutch texts, 92% of all word tokens have default stress (Schiller et al. 1997). In other words, for the majority of words produced, speakers of Dutch (and probably speakers of English) retrieve no independent metrical word frame.

Does this undermine MacNeilage's main thesis about the speech producing mechanism? It certainly contradicts his claim that there is good evidence for syllable structure information being represented in the mental lexicon. But that claim is not an indispensible ingredient of MacNeilage's theory. The core point is that there is a basic syllable cycle in the speech production system. That basic syllable cycle can run without the retrieval of stored syllable frames. Syllabification is probably a late process in phonological encoding. Spelled-out (retrieved) segments are incrementally chunked into syllabic units, because repetitive "legal" syllabic structure is the exclusive target structure at this level of phonological word encoding.

## Speech evolution: Let barking dogs sleep

Philip Lieberman

*Department of Cognitive and Linguistic Sciences, Brown University, Providence, RI 02912-1978.* **philip_lieberman@brown.edu**

**Abstract:** Many animals, including dogs, produce vocal signals in which their mouths open and close producing "syllables." In contrast, the vocal signals of species other than humans are tied to emotional states. The Broca's-Wernicke's area model of the brain bases of language is wrong.



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

I cannot quarrel with MacNeilage's argument that human language evolved by means of Darwinian processes. Moreover, MacNeilage is to be commended for proposing an explicit hypothesis consistent with Darwinian theory, rather than a repeated mantra to the effect that language and a "universal Grammar" must have evolved (Pinker & Bloom 1990). It is also evident, as MacNeilage and many other scholars have noted, that human speech differs qualitatively from the vocal communication of other species.

It is not clear, however, that MacNeilage is right that the "main difference between speech and other mammalian call systems," is that the human "vocal tract alternates more or less regularly between a relatively open and a relatively closed configuration (open for vowels and closed for consonants)," (sect. 2.2) and that this alternation is the basis of the syllable. Acoustic analysis of primate vocalizations, for example, shows formant frequency transitions similar to those that convey the consonant-vowel distinctions of human speech (Lieberman 1968; 1984). Similar alternations involving mandibular movement typify one common animal communication, the barking of dogs. If such articulatory alternations were the touchstone of human speech, dogs would be able to talk. Moreover, although syllables generally have the closed-open, consonant-vowel structure that MacNeilage notes, syllables do not have to conform to this pattern. For example, consider the English sentence, "How are you?" Other examples can be found in many other languages. The syllable is perhaps best regarded as a minimal unit for speech encoding.

MacNeilage's scenario for the evolutionary significance and modification of Broca's area appears to be based on the Lichtheim (1885) model of the neural bases of human language, in which Broca's and Wernicke's areas are connected by a putative cortical pathway. It is evident, however, that the Lichtheim model is incorrect. Permanent aphasia never occurs in the absence of subcortical damage (Stuss & Benson 1986). Furthermore, aphasia and aphasia-like phenomena occur when subcortical basal ganglion structures are damaged, or disease processes affect their function. (Alexander et al. 1987; Lieberman 1991; 1998; Lieberman et al. 1992; Naeser et al. 1982; Natsopoulos et al. 1993; Pickett et al. 1998). The evolution of human speech and language may have involved Broca's area, but as Mesulam (1990) has noted, the neural substrate of human language appears to be a distributed network that clearly incorporates subcortical components as well as cortical regions; recent studies confirm this view (Lieberman, in press). It is perhaps time to go beyond Lichtheim's (1885) model.

Hence, MacNeilage's proposals concerning the transfer of motor control from the oscillatory control of chewing to speech in Broca's area are probably beside the point. Similar open-closed sequences typify animal communication, including the sustained barks of dogs; the syllable structures that occur in human languages do not always conform to this formula, and the probable neural bases of human speech transcend the Lichtheim (1885) model.

What does appear to be unique to human speech is that we can utter sounds that are not "bound" to particular emotions or affective states (Lieberman 1994). Certain primate vocalizations are to a degree independent of emotion or arousal; they may represent the initial stages of the neural evolution that allows us to talk and probably allows us to think (Lieberman 1995; 1998). However, the magnitude of the quantitative distinction yields a qualitative distinction.

(For the record, in 1940 Roman Jakobson noted the correspondences between the frequency with which particular syllable structures occurred in different languages and the ontogenetic development of speech; Johannes Muller [1848] also noted this correspondence.)

## A curiously ubiquitous articulatory movement

Björn Lindblom

*Department of Linguistics, Stockholm University, S-10691 Stockholm, Sweden, and Department of Linguistics, University of Texas at Austin, Austin, TX 78712-1196.* **lindblom@ling.su.se**
**www.ling.su.se/staff/lindblom/lindblom.html**

**Abstract:** The frame/content theory justifiably makes tinkering an important explanatory principle. However, tinkering is linked to the accidental and, if completely decoupled from functional constraints, it could potentially play the role of an "idiosyncracy generator," thus offering a sort of "evolutionary" alibi for the Chomskyan paradigm – the approach to language that MacNeilage most emphatically rejects. To block that line of reasoning, it should be made clear that evolutionary opportunism always operates within the constraints of selection.

***A remarkable fact in need of explanation.*** It is indeed curious that all spoken utterances are superimposed on a "continual rhythmic alternation between an open and a closed mouth (a frame)" (target article, sect. 1). MacNeilage is certainly right in observing that this movement is pervasive in spoken language and that we need to find out why. For example, with very few exceptions, the favored syllable structures of the world's languages are all variants of a sequence of close-open-close articulations. Phoneticians and linguists have known this for a long time (cf. de Saussure's description of the "*syllable*" [Malmberg 1960] as an opening-closing movement). But they have not taken it very far theoretically. Consequently, the observation remains a descriptive statement without an explanation. Why should the complex edifice of Language rise from this simple cyclical movement? Why is it never abandoned? Why are many other movement patterns not (also) used? The surprising fact is that they are not.

***Physical constraints on mandibular movement.*** The jaw movements of speech put into play a mechanism shaped by vegetative functions such as chewing. Mechanically, the mandible has mass, elasticity, and damping. The associated neural structures (e.g., central pattern generators) have certain "delay line" characteristics that presumably arose by adaptation to the physics of the system and by imposing constraints on how smoothly and quickly the system can switch between lowering and elevating the jaw. Together, these factors create a virtual oscillator with over-damped (sluggish) response characteristics and with preferred rates of open-close alternation ("resonance" characteristics). The system can be compared to a swing. Keeping it going at its own preferred rate (natural frequency) requires only a small, appropriately timed push, but to activate it at other rates requires more than a push and entails greater energetic costs. Paraphrasing frame/content theory (F/C), we can say that there is an open/close alternation in speech because prehistory made it available and evolution opportunistically made use of it. Importing the "frame" was a phylogenetically small step. However, in view of the preceding analysis, it seems important to add that this act of tinkering was probably also promoted by the fact that the behavior was energetically cheap.

***Clues from mouth sounds and vegetative phenomena.*** This supplementary criterion may have been a crucial aspect of the selection process. Speech makes fastidious use of the space available in principle for sound production. The world's speech sounds form a small set compared with the total capacity for vocal gesture and sound. They seem to be located at the low-key end of the full range of possibilities, whereas vegetative behaviors, including "*ingestive cyclicities,*" demand much higher efforts. This can be seen when swallowing occurs between speech samples during an electromyographic (EMG) experiment. With the gain of the signals adjusted for speech, swallowing will throw the signal way out of bounds. As many dentists will readily tell us, mastication also requires very high force levels. It recruits powerful muscles such as the temporalis and the masseter, not needed for speaking but absolutely crucial to chewing and cracking nuts and bones open.

Nonspeech activities demonstrate more fully the entire capac-



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

ity of the system. In contrast, speech comes across as a physiological pianissimo. Like many other movements, vocal gestures are shaped by their purpose (Granit 1977).

***Evolution is also a miser.*** Was speech assembled from preexisting vegetative behaviors? Was the ingestion-related mandibular oscillation simply taken over intact by what was to become speech? No, it is likely that its adoption occurred with significant modifications brought about by the demands of its new use. If a behavior is energetically extravagant, there is likely to be a reason for its staying extravagant. If there is no such reason, the selection process will prune it and reduce costs.

The point made here is illustrated by the following anecdote. Henry Ford is reported to have asked whether scrapped cars came back with any spare parts still usable. The answer was that there was indeed something that never wore out. Ford immediately had the part replaced by one of inferior quality. No extravaganza unless it serves a purpose (Humphrey 1986).

***A caveat concerning tinkering.*** Why is the "frame" so pervasive in the organization of speech movements? Why are other types of movement patterns not used instead?

It is to the considerable credit of F/C theory to have (1) put those questions into the scientific spotlight and (2) launched a search for an explanatory answer by tracing a *continuous* path from precursors to present-day speech behavior. By so doing, the F/C research program makes an important contribution toward replacing the twentieth century tradition of "descriptive structuralism" with a type of language study that is more firmly anchored in biology and therefore able to offer better prospects for deepening our understanding of language.

F/C justifies making tinkering an important explanatory principle. Tinkering, however, is linked to the accidental. Consequently, if completely unconstrained and decoupled from functional principles, it provides a mechanism for generating arbitrary idiosyncracies, Rube Goldberg contraptions, hopeful monsters, and autonomous structures such as Universal Grammar. Clearly, evolutionary tinkering does not work in that way. What needs to be underscored in the F/C scenario, then, is that tinkering never works alone. It always occurs within margins set by constraints.

## Is speech just chewing the fat?

James P. Lund

*Faculty of Dentistry, McGill University, Montréal, Québec, Canada, H3A 2B2.* **lund@medcor.mcgill.ca**

**Abstract:** It is likely that the system controlling speech has evolved from the one that controls feeding. However, the idea that frames and content are programmed independently by two different cortical areas is not plausible. Models of the speech control system must also take into account the need to coordinate the respiratory, laryngeal, and articulatory musculature.

McNeilage suggests (sect. 4.2) that speech makes use of the same brainstem central pattern generator (CPG) as ingestion. If this is true, it is certainly not the first time that "tinkering" with this system has gone on during the evolution of vertebrates. Just as the basic circuits controlling locomotion were present in the spinal cord of early vertebrates (Grillner 1985), so were the circuits controlling ingestion and respiration. Long before the appearance of jaws and lungs, lampreys had trigeminal, vagal, facial, and hypoglossal motor nuclei (Nieuwenhuys 1972) that were eventually coopted for the control of new structures. We have found that the interneurons controlling trigeminal motoneurons are located in similar sites in the brainstem of lampreys and mammals (Huard et al. 1995), and that some of these interneurons pattern ingestion in lampreys (Petropoulos et al. 1997). We presume that they have retained this role during the evolution of mammals.

Mastication is cyclical, but the movements vary greatly with food type and between cycles in response to feedback and central drive (Lund 1991). Unlike locomotion, basic patterns of mastication are represented in a site-specific manner in the lateral sensorimotor cortex of lower species (Bremer 1923; Lund et al. 1984), and in areas 6, 4, and adjacent pre- and post-central sites of monkeys and humans (Beevor & Horsley 1894; Hung et al. 1989; Lund & Lamarre 1974). The face area of the supplementary motor cortex on the medial wall is also active during ingesting (Picard & Strick 1997).

It is unlikely, however, that the various patterns are elaborated in the cortex because they can be evoked by stimulating descending tracts and sensory inputs in decerebrate animals (Bremer 1923; Schwartz & Lund 1995). Instead, these cortical areas appear to act through a brainstem Central Pattern Generator (CPG) made up of subsets of neurons that participate in one or more of these patterns but not all of them (Westberg et al. 1995).

Although I agree with MacNeilage that it is very likely that this highly flexible control system has been coopted for the control of speech during the evolution of mankind, he has neglected several major points:

(1) Monosynaptic connections between the lateral regions of the cortex and the trigeminal, facial, and hypoglossal nuclei of humans appear in the higher apes (Kuypers 1958). If they are not participating in the control of facio-visual and verbal communication, why did they evolve?

(2) Speech is the product of the coordinated activity of articulatory and respiratory motor systems, but masticatory and respiratory rhythms are not inherently phase coupled (McFarland & Lund 1995). How is this higher level of control exerted during speech if not by the cortex?

(3) Finally, MacNeilage really has no evidence that frames and their content are controlled by independent circuits. The medial cortex plays a crucial role in the production of the vocal repertoire of nonhuman mammals, which, according to MacNeilage, is characterised by an absence of rhythmic modulation (sect. 2.2). One of the major premises of his Frame/Content theory is that, during evolution, the medial cortex took on the production of opening and closing of the mouth (the frames), leaving the "content" to the lateral cortical areas that also participate in ingestion. This seems both illogical and unnecessary, because frame and content are *not* independent in ingestion. Both are generated by the brainstem CPG when it is driven by the lateral cortex, and both are usually modified in parallel by peripheral or central inputs. Contrary to the statement in section 3.2, splitting frames and content is no "natural division of labor." Indeed, there really is nothing to divide.

## Frame dominance: A developmental phenomenon?

Lorraine McCune

*Department of Educational Psychology, Graduate School of Education, Rutgers University, New Brunswick, NJ 08903.* **mccune@rci.rutgers.edu**

**Abstract:** Developmental aspects of the frame/content perspective are explored in relation to (1) transitions in early language acquisition, (2) possible differential neurological control for babbling and early and later speech, and (3) development of word production templates in precocious early speakers. Proportionally high frequency of bilabial stops in early stable words versus babble offers advantages for afferent monitoring and supporting "frame dominance."

The development of speech motor control as a dynamic system beginning with the mandibular central pattern generator is theoretically attractive. Supportive findings have thus far been limited primarily to studies by MacNeilage and associates (e.g., Davis & MacNeilage 1995). Vihman (1992) tested consonant/vowel associations in 23 children across the 3-month period just prior to the emergence of words (9–12 months of age) and found little support



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

for the predictions. These differences may be the result of methodological factors, but neither transcription nor acoustic analysis methodologies provide a strong test of these motor hypotheses, particularly given the known capacity of humans to achieve similar vocal output by varying strategies.

Given the lingual movements apparent in suckling and in the integration of respiratory, laryngeal, and oral movements when eating from a spoon, movement that most infants demonstrate by 6 months of age (the lower limit of the babbling period), why might limitations on such integration rule speech sounds? Although learned movements such as those specific to speech may share neural mechanisms with innate patterns, "different motor control may be required when [the same] movements are used in meaningful speech" (Barlowe & Farley 1989, p. 85).

Recent neurological research demonstrates different cortical processing for language comprehension in children differing in language level and suggests that experience with language may influence neurological structure (Mills 1994). Anatomical findings more directly related to production demonstrate initial bilateral advantages in dendritic length and branching for cortical motor areas supporting orofacial, laryngeal, and pharyngeal mechanisms that show rapid increases at 5–6 months of age, the period of transition to prelinguistic babbling, in comparison to Broca's area and the right hemisphere analogue. A shift to a left Broca's area advantage for distal dendritic segments begins at 12–15 months, and is more fully apparent at 24–36 months, spanning the period of initial language learning (Scheibel 1993). The frame/content (F/C) developmental proposals need to be refined in relation to potential differences in neurological control for earlier and later speech versus babbling.

What developmental steps can be predicted from F/C within the early phases of speech? During the prelinguistic period laryngeally based vocalizations occur in relation to metabolic demands (grunts; McCune et al. 1996) and affective state (e.g., distress cries, comfort sounds, squeals of joy, laughter) until, in conjunction with the development of other rhythmicities (Thelen 1981), rhythmic jaw movements break the vocalic airstream and babbling begins. As early as 9 months of age grunts accompany focal attention (possibly because of mental effort) and early context-limited words partaking of the babbling repertoire are noticed. It seems likely that frame dominance would govern this early period.

McCune et al. (1996) found that "early bloomers" who produced communicative grunts by 13 and 15 months of age, followed by their first referential words and a spurt in lexical production by 16 months, differed phonetically from later language producers. Children making this transition early were distinguished from their peers by demonstrating larger numbers of vocal motor schemes for consonant production, including the stable capacity to produce the bilabial stop (p/b) in babble beginning at 9–10 months (McCune & Vihman 1987; submitted). As reported in many studies, (p/b) dominated their stable word production (in this study it was nearly 50%) in contrast with t/d, equally available as a vocal motor scheme in babbling, but occurring in only about 10% of stable words.

An initial (p/b) advantage supports frame/content as the basis for pure frames and further links with the potential significance of such visible and audible gestures as the "lipsmack," common to other primates as well as humans. (Blown any kisses lately?) In addition, neurological control of (p/b) differs from that for other consonant segments. Studies of motor equivalence and compensatory articulation imply that an efference copy accompanies motor commands, allowing afferent systems to monitor goal achievement continually (Evarts 1982); (p/b) also provides special opportunities for afferent feedback. Bilabial closure for (p/b) is achieved primarily by lower lip elevation supported by the innately available jaw movements forming the foundation of frame/content. A strong proprioceptive afferent signal based on cutaneous pressure across total lip surface guides the production of (p/b), in contrast with other consonants (e.g., [t/d]), where the articulatory targets are finer and where afference from oral structures relies on muscle spindles or Golgi tendon organs that do not exist in the lips. This broad and direct feedback may provide a sensorimotor advantage for participation of this consonant in speech, as well as providing an initial afferent basis for learning linguistic oral motor control.

At a more advanced phase of phonetic control some precocious early speakers demonstrate lexical production patterns (templates) that come to dominate their productions, deforming words of the ambient language to suit their individual production repertoires (Vihman & Velleman 1987; Vihman et al. 1994). These templates can be considered as idiosyncratically organized "starter" frames that launch the children on linguistic patterns that will eventually converge on the flexible ability to produce the ambient language in a manner similar to adults, demonstrating a possible developmental trajectory within frame dominance. Given the lingual integration and consistent intonation of the most complex of these templates, the integration of respiratory, laryngeal, and beginning articulatory control seems likely at this point.

## A multi-modal, emergent view of the development of syllables in early phonology

Lise Menn

*Department of Linguistics and Institute for Cognitive Science, University of Colorado, Boulder, CO 80309-0295.* **lise.menn@colorado.edu**
**www.colorado.edu/linguistics/faculty/lmenn/Home.html**

**Abstract:** A narrow focus on the jaw (or on motor generators) does not account for individual and language-specific differences in babbling and early speech. Furthermore, data from Yoshinaga-Itano's laboratory support earlier findings that show glottal rather than oral stops in deaf infants' babbling: audition is crucial for developing normal syllables.

Legitimate excitement about the power of one contributing factor may lead people to dismiss factors that are equally important. MacNeilage has a good grip on the elephant's trunk, but he appears to be ignoring reports from colleagues who have been patting its flanks and pulling its tail. The syllable is not adequately described ontogenetically in terms of jaw motion alone; furthermore, segment coarticulation apparently develops in at least two ways: (a) by differentiation from syllable-length gestures in which the segmental targets may not be represented as such, as MacNeilage and associates have shown, and (b) by integration of poorly coarticulated gestures (Hawkins 1984). To cite Kent (1993, p. 120): "Motor control is very much like the development of skilled movements generally: The child gradually increases in speed, precision, and – at least in some respects – the degree of anticipation in a motor sequence." Nor is the extent of coarticulation fixed in adult speech. De Jong et al. (1993) have shown that stressed syllables have less coarticulation than unstressed syllables; stressed syllables are hyperarticulated.

Even those children who start from long units need not take the syllable as a base; there is also "a completely different type of organization . . . in which the critical organizational unit seems to be an autosegmental "melody" . . . which ranges from a single syllable nucleus to two syllables" (Vihman 1993, p. 165). There is more than one way to approach adult targets. The fact that some developmental patterns are more common than others is a phenomenon that demands explanation, but such explanations cannot be of a form that would preclude the existence of the less common patterns.

The syllable is constructed (in normally hearing and articulating infants) from at least three sensory ingredients besides rhythmic motor jaw opening and closing: air-passage opening/constriction proprioception, air-flow/blockage sensation, and hearing the sound amplitude change when phonation accompanies free versus impeded airflow. The syllable is thus emergent (self-organizing) from at least four prelinguistic, internally generated sensorimotor



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

sources, as well as from the phonotactic patterns of the ambient language, which children begin to apprehend between 6 and 9 months of age (Jusczyk et al. 1993). (In addition, temporal modulation of the syllable frame is not entirely a matter of the segments inserted into it. Appeals to readers' supermarket observations notwithstanding, language-specific departures from speech timing, such as phrase-final lengthening, emerge in babbling [Levitt & Wang 1991], and cannot be ascribed to the segments inserted in a frame.) Even within the motor arena, jaw oscillation is only part of early oral play (Stark 1980, p. 85): "The tongue is protruded and retracted, the lips rounded, the mouth silently opened and closed." Articulatory Phonology (Browman & Goldstein 1992; 1995), though limited to a motor focus, can capture the emergent character of the syllable in a way that other approaches cannot.

Oral stops, far from being motor-driven, rarely emerge without auditory feedback and perception of the ambient language (Oller & Eilers 1988; Stoel-Gammon & Otomo 1986). The babbling/early speech of 30 deaf children of hearing parents, recently analyzed by Valerie Wallace in Christine Yoshinaga-Itano's presentation to research group at the University of Colorado, October 1997, support the conclusions of the earlier studies: mildly to profoundly hearing-impaired babies between the ages of 6 months and 1 year are conspicuously different from those with normal hearing in having almost *no* jaw-based rhythmic babble. When they do produce syllables, the consonant is almost always glottal (stop or fricative).

Only 8 of the 30 children had oral stops; in those 8, oral stops were found in only 8% of their productions (on average; the maximum was 22%). Although 10 of the 30 children were evaluated at age 6 months, and could plausibly have gone on to develop oral-stop babble later, only 4 out of 10 who were evaluated at about age 12 months had any oral stops.

Returning to Kent (1993, p. 120): "Motor activity is governed by goals, tasks, or objectives. A motor score as a kind of prescription for motor activation needs itself to be informed." An integration of Articulatory Phonology with equally well-developed acoustic/auditory phonetics should provide a proper basis for understanding the ontogeny (and contemplating the phylogeny) of speech.

ACKNOWLEDGMENTS

I am grateful for discussions with Valerie Wallace and other members of Yoshinaga-Itano's research group (supported by NIH contract NO1-DC-4-2141), and for materials and responses to questions from Louis Goldstein.

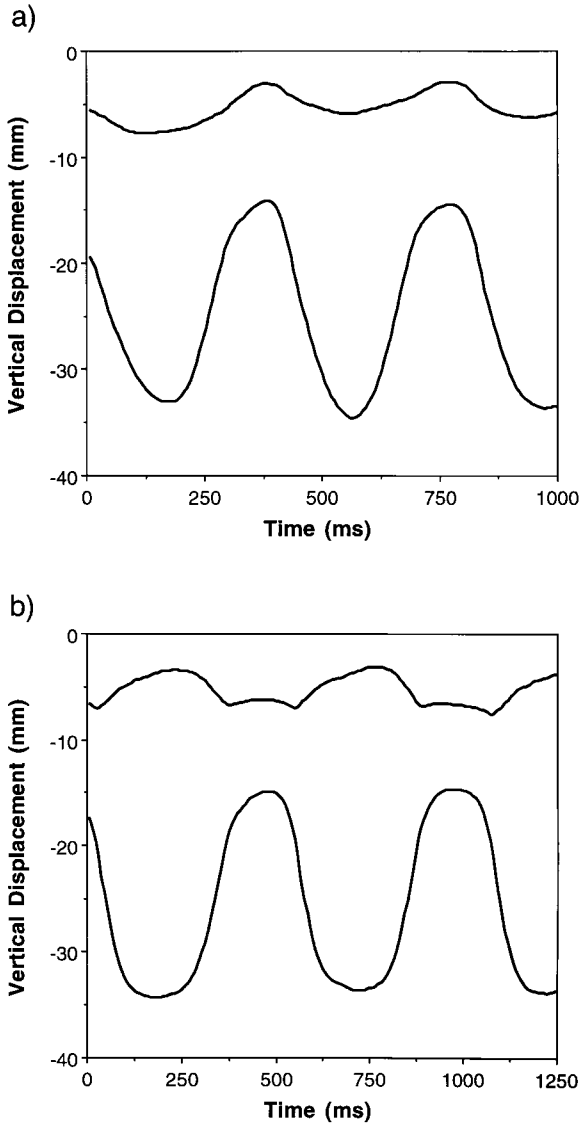



Figure 1 (Munhall & Jones). Vertical displacement of upper lip (top trace in each figure) and lower lip (bottom trace) as a function of time during repetitive production of /bababa/ by (a) an 8-month-old and (b) an adult.

## Articulatory evidence for syllabic structure

K. G. Munhall and J. A. Jones

*Department of Psychology, Queen's University, Kingston, Ontario, Canada K7L 3N6.* **munhallk:psyc.queensu.ca**
**www.psyc.queensu.ca/faculty/munhall/munhall.html**

**Abstract:** Because the evolution of speech production is beyond our expertise (and perhaps beyond everyone's expertise) we restrict our comments to areas in which data actually exist. We provide articulatory evidence consistent with the claims made about syllable structure in adult speech and infant babbling, but we also voice some disagreement about speech errors and the typing data.

Over the past few years we have made a number of attempts to study the kinematics of infant babbling. We offer Figure 1a as anecdotal evidence from an 8-month-old baby babbling. The data were collected using OPTOTRAK with infrared emitting diodes attached at the midline of the upper and lower lips close to the vermilion border. The most important point to note about this figure is that the movement does not appear to involve the upper lip actively. The only upper lip movement occurs in phase with the lower lip motion and is presumably caused by the lower lip forces pushing the upper lip upward after contact. This pattern is consistent with the proposal that initial babbling primarily involves mandibular motion; the lower lip rides on the jaw and deforms and upper lip on contact. In contrast, Figure 1b shows an adult producing the same sequence. As can be seen, the upper and lower lips both produce opening gestures. In addition, some deformation of the upper lip can be observed at the lower lip's upward displacement peak. As MacNeilage suggests, early babbling could be produced by a simple "frame" style of speech motor control using a repetitive jaw cycle.

In adults we have been pursuing studies of spontaneous articulation changes at fast speaking rates (Jones et al. 1997). The subjects repeat bisyllables at increasing speaking rate and we examine the changes that occur at the fastest rates (cf. Stetson 1951/1988). One particular observation is relevant to the framework proposed by MacNeilage and others. When subjects repeat words such as the surname "Thompson"[1] at increasing rates, they frequently switch to saying "Thompsom" at the fastest rates.[2] This harmonizing of the syllable codas is remarkable because the im-


https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

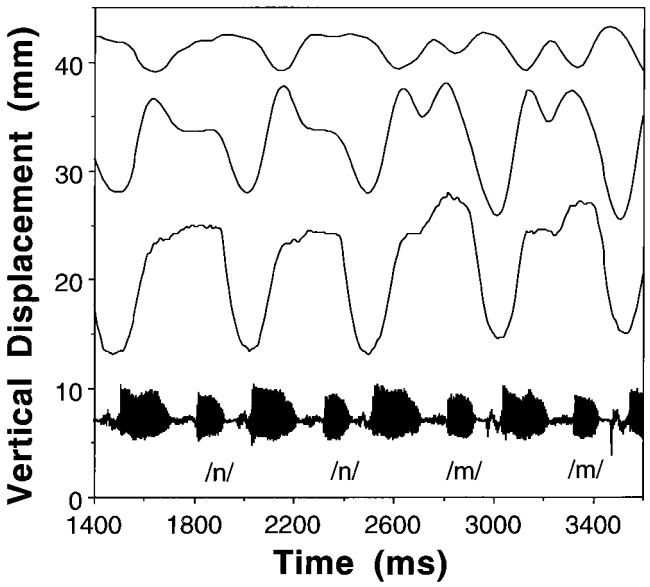



Figure 2 (Munhall & Jones). Vertical displacements of the upper lip (top trace), lower lip (second trace), tongue tip (third trace), and acoustics as a function of time. (Note that an arbitrary reference frame has been used for display purposes.)

mediately surrounding consonants in the series, /s,t/, share place of articulation with /n/. Yet, the /n/switches to a bilabial to be in harmony with the closest coda. We view this phenomenon as strong evidence for the psychological reality of syllabic structure in real-time speech production processes. Figure 2 shows X-ray microbeam data of vertical movement of the upper lip, lower lip, and tongue tip during a trial of "Thompson" repetition at increasing rates. As can be seen, the tongue tip and lips show a transition as the rate increases. The lips spontaneously begin to make a second /m/ within the word at precisely the point in the utterance at which the tongue tip behavior changes. No major changes are noted in the mandible movement.

It is clear that the change observed in our data is influenced by the syllable structure outlined by MacNeilage and by others elsewhere; it is not clear, however, that the best characterization of this transition is a segment switch rather than a place of articulation feature switch. We disagree with MacNeilage's assertion that the speech error data show no evidence for subsegmental patterns. There is considerable evidence that people have difficulty detecting feature-sized errors, and transcription bias is well documented (e.g., Buckingham & Yule 1987; Itoh & Sasanuma 1984; Kent & Rosenbek 1983).

Finally, we would like to comment on MacNeilage's claim about typing and errors. All movements are organized with respect to a number of frames of reference. The coordinate frames for speech include the spatial framework of the vocal tract and we believe, along with MacNeilage, that syllabic structure coordinates are fundamental to articulation. To say that typing does not share these exact coordinate frames should not be controversial. For example, one would not expect vowels and consonants to have privileged status in transcription because these categories are uniquely defined with reference to the vocal tract. The coordinate frames for typing are obviously different and include the spatial layout of the keyboard and the coordination of the fingers and of the two hands with respect to this keyboard space. MacNeilage is right to say that the largest number of typing errors are adjacent keys on the keyboard, but we think he underestimates the number of similar errors in articulation. Momentary sloppiness in speech and subtle distortions in articulation are widespread in speech but are seldom represented in speech error counts and databases. Furthermore, there is structure to errors in typing that does reflect the other coordinates of motor organization of the hands for this task. In a study of mirror-image movements in typing (Munhall & Ostry 1983) we found that sequential mirror-image movements had longer between-key intervals than the same key typed following a key that was not its mirror-image. In addition, mirror image errors between the hands occurred far more frequently than chance, given the frequency of the letters in English. This was true of data typed on a standard QWERTY keyboard and also of data compiled from typists using a Dvorak keyboard. Thus, the errors were not caused by the statistics of English letter sequences, but were influenced instead by the spatial coordinate system with respect to which typing is organized.

ACKNOWLEDGMENT

This work is supported by NSERC and NIH grant DC-00594.

NOTES

**1.** The "p" in the orthography is irrelevant because the same effect is observed for "Thomson." Both spellings have epenthetic /p/ at the fastest rates.

**2.** The reader can replicate this finding by saying "Thompson" at a normal rate and then repeating it. With each repetition say the word faster until the fastest rate is reached. At some point in the sequence, most speakers switch to "Thompsom." As a control, say "Thonson" in the same manner. No switch is observed for this utterance.

## Content first, frame later

John J. Ohala

*Department of Linguistics, University of California, Berkeley, CA 94720.*
**ohala@cogsci.berkeley.edu trill.berkeley.edu/users/ohala**

**Abstract:** There is not enough reason to believe that syllables are primary in speech and evolved from the cyclic movements of chewing. There are many differences between chewing and speech and it is equally plausible that what is primary in speech is a succession of auditorily robust modulations of various acoustic parameters (amplitude, periodicity, spectrum, pitch); syllables could have evolved from this.

MacNeilage's idea that chewing might be a precursor to speech is not completely new. It was advocated a half century ago (1951) by Weiss and was the basis for the "voiced chewing" therapy he prescribed for stuttering and other speech pathologies. But except for Froeschels (1951), there seems to have been little support for this idea. MacNeilage has elaborated this notion and taken it much farther, but I do not think the evidence currently available is sufficient to support it.

There is a superficial resemblance between chewing and speech, but in many details the two are quite dissimilar. Chewing necessarily involves a significant lateral movement of the jaw that is irrelevant to and largely absent in speech. Chewing requires no auditory feedback, whereas auditory feedback is essential to speech. Typically during chewing (as distinct from swallowing) the soft palate is lowered for breathing; this is not true for speech, where the soft palate lowers only during nasal sounds like [m] or [n]. It might be claimed that these differences fall under the "modification" part of the Darwinian notion of "descent with modification," but as far as I know there is as yet no rigorous way to constrain evolutionary scenarios of this sort. The problem is analogous to what originally faced etymology, the discipline that traces the evolution of words. Until rigorous methods were established by linguists in the early 19th century, all sorts of fanciful word histories were offered.

In the frame/content theory, the syllable or the opening-closing of the vocal tract is the frame that is evolutionarily primary, and the individual speech sounds or phones are the content fitting in that frame, came later. It is equally possible and plausible, how-



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

ever, to advocate just the opposite: phones were primary and syllables are epiphenomenal developments from them. MacNeilage says that in human speech, "the vocal tract alternates more or less regularly between a relatively open and a relatively closed configuration," (sect. 2.2), but there are many types of syllables where this is not true, notably those with high close vowels such as [i u]. Here the jaw does not open very much and there may be no more mouth opening (and perhaps less) than in some consonants, for example, [j w] (as in the initial sounds in "you" and "we") and the glottal consonants [h] and glottal stop. Many languages, English included, also have some syllabic consonants, such as the second syllable in "button," where there is no mouth opening-closing movement. One could dismiss these as being somehow inconsequential exceptions or modifications of the more basic open-closed cycle, but there is really nothing marginal about such syllables.

What all speech does have, and this includes MacNeilage's example of the word "tomato" (sect. 2.2) as well as "button" and complex syllables like English "strengths," is *sequences of acoustic modulations.* The amplitude, periodicity, spectrum, and pitch are modulated, that is, varied, to create the kinds of differences in sounds that the listener requires (MacNeilage covers this briefly in sect. 5.5). I regard such acoustic modulation as the primary physical characteristic of speech (Ohala 1995; Ohala & Kawasaki-Fukumori 1997). How does the syllable evolve from this?

Imagine that one starts from a closed vocal tract in making, say, the bilabial stop [b]. What can one do from that configuration to make an auditorily detectable modulation? The answer is that one can open one's lips but keep a partial closure in the vocal tract with the tongue touching one side of the hard palate, making, say, a lateral [l]. What then? One can release the lateral closure and open the mouth further to the vowel [a]. With the vocal tract maximally open there are few other opportunities to make another acoustic modulation except by beginning to close the vocal tract, say, to a slight palatal constriction to produce the palatal glide [j]. And so on.

At intermediate degrees of vocal tract closure one has the option of making the next gesture a more open vocal tract (in the oral cavity, the velopharyngeal area, etc.), but from a maximal closure one can move only to a more open tract and from a maximally open one, to a less open one. From this viewpoint what are called "syllables" are just epiphenomenal consequences of the necessity of making a succession of auditorily robust modulations in one or more acoustic parameters. It may be that after the evolution of full, articulate speech, a frame-like function of the syllable was imposed, and this may account for the compelling speech error data and tip-of-the tongue data, but this does not necessarily imply that syllables were there from the start.

One final quibble: MacNeilage states that "infants are born with the ability to phonate, which involves the cooperation between the respiratory and phonatory systems" (sect. 5.2). We would do well not to exaggerate the degree of cooperation between the lungs and the vocal organs in newborns: as is well known, the majority of newborns have cries that include a short inspiratory "coda" to a long expiratory cry (Grau et al. 1995; this is heard as something that sounds a bit like a "hiccup"). This occurs when the vocal cords remain in the phonating configuration while the infant breathes in. This feature eventually disappears and cries are almost entirely superimposed on expirations. The point is that this discoordination between lungs and the vocal organs is present at birth.

## Out of the mouths of babes . . . and beaks of birds? A broader interpretation of the frame/content theory for the evolution of speech production

Irene M. Pepperberg

*Department of Ecology and Evolutionary Biology and Department of Psychology, Program in Neurosciences, University of Arizona, Tucson, AZ 85721.* **impepper@u.arizona.edu**

**Abstract:** Much of the material MacNeilage cites to support his frame/content theory for the evolution of speech production in humans is not unique to mammals. Parallels can be drawn for comparable evolution of vocal flexibility (specifically the reproduction of human speech) in birds. I describe several such parallels and conclude that MacNeilage's hypotheses may have broader application than he envisioned.

MacNeilage presents a fresh approach to a difficult problem. He proposes that mammalian speech production evolved from control systems used in ingestive behavior, with brain function separating into areas responsible for motor control and vocal learning. I do not disagree with MacNeilage, but contend that his hypothesis also applies to birds.

MacNeilage argues (sect. 2.1) that the two-tube vocal tract is unique to hominids; in recent studies (Patterson et al. 1997; Patterson & Pepperberg 1998; Warren et al. 1996), Grey parrots' (*Psittacus erithacus*) production of human vowels and consonants can be explained and modeled only by positing the existence of a two-tube vocal tract. Parrots, too, produce "rapid and highly variegated sound sequences in syllabic packages." Replication/use of human speech is not a natural psittacine behavior; maybe this ability evolved merely for vocal flexibility.

MacNeilage states (sect. 2.2, para. 2) that open/closed alternation of the articulatory system during vocalization is a "defining characteristic" of human speech. In Patterson et al.'s (1997) model of Grey parrot speech, glottal opening/closing, for example, is critical for consonant production. Beak opening/closing also affects speech (Warren et al. 1996).

MacNeilage supports his hypothesis with speech error data (sect. 3.1), in which replacements follow rules. Little information exists on psittacine speech errors; however, English "sound play" of Grey parrots follows comparable rules (Pepperberg et al. 1991, unpublished data): We find progressions like "grey," "grain," "chain," and "cane," but not "achn."

MacNeilage supports a slot/segment hypothesis (sect. 3.2). Very preliminary data on Grey parrots (Neal 1996) suggest similar behavior: at the earliest stages of label acquisition, birds produce only a label's vowels, but they do appear to reserve spaces in timing for missing consonants.

Given MacNeilage's emphasis on syllables, segments, and occasionally phonemes, note that Grey parrots produce *meaningful* minimal pairs (e.g., *tea, pea*): They understand that requesting the former provides a sip of liquid and the latter, a vegetable (Pepperberg 1990).

MacNeilage derives human speech from mammalian ingestive behavior (sect. 4.2); Homberger (1986) suggests that the flexibility of the Grey parrot lingual apparatus, used in vocal production, arose as an ingestive adaptation. The statement (sect. 4.3) about mammalian use of mandibles for manipulating objects is also noteworthy: the clumsiness of parrot claws compared to primate hands for object manipulation, coupled with parrots' abilities to use their beaks and tongues to achieve much of what nonhuman primates accomplish with digits, support speech evolution in parrots even more strongly than in humans.

MacNeilage proposes how ingestive cyclicities might get into the mammalian communicative repertoire (sect. 4.4); one can similarly judge avian juvenile begging calls and motions. Budgerigar (*Melopsittacus undulatus*) begging actually develops into adult contact calls (Brittan-Powell et al. 1997).

MacNeilage notes perceptual benefits of producing varied tran-



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

sients at high rates (sect. 5.5), and how this consonant-linked capacity could be an important hominid-specific communication development. Note that Grey parrots and budgerigars can also produce *and* distinguish among human consonants (Dooling et al. 1989; 1995; Patterson & Pepperberg 1998).

MacNeilage discusses how brain structures for imitation and motor control seem tied to speech learnability (sect. 6.6). But psittacine imitative capacities and songbirds' abilities to learn vocalizations are correlated with specific brain structures; and avian brain structures responsible for vocal tract control and vocal learning likewise seem integrated (Nottebohm 1980; deVoogd et al. 1993; Durand et al. 1997; Striedter 1994). Physical mimicry may also exist in Grey parrots (e.g., Moore 1992), although the actions described are not novel.

MacNeilage's finale (sect. 7) is most intriguing. At present, avian fMRIs are not possible; other techniques also have drawbacks. Still, experiments may someday determine whether MacNeilage's hypothesis for the evolution of human speech holds for birds. It is interesting that, apropos of MacNeilage's final points on laterality, a parallel may exist with foot-dominance/repertoire-size in Grey parrots (Snyder & Harris 1997).

In sum, *language* may be unique to humans, but *speech capacity* is not: MacNeilage's hypothesis may have implications beyond mammals.

## On mandibular oscillation as a source of variation in infant vocalizations

Jörg Peters

*Institut für Germanistik, Universität Potsdam, 14415 Potsdam, Germany*
**jpeters@rz.uni-potsdam.de**

**Abstract:** The target article raises the question of whether transcription-based evidence is sufficient to support assumptions relating to patterns of mandibular activity in young children. Studies on the perception of both adult and infant speech indicate that the argument needs to be reexamined on the basis of acoustic and articulatory data.

In recent years, MacNeilage and his colleagues have argued convincingly that there are a number of preferences in the co-occurrence of phonetic segments in infant babbling. These preferences are seen as evidence that phonetic variation in early utterances is primarily caused by variation in mandibular oscillation. The bulk of the evidence advanced in support of this argument has been based on an analysis of perceptual features of babbling. It accordingly seems appropriate to ask whether indirect evidence of this kind is sufficient to support hypotheses relating to patterns of mandibular activity.

The problems inherent in adopting such an approach are well exemplified by hypothesis (5), outlined in section 5.2, in which MacNeilage suggests that there will be relatively more intersyllabic changes in tongue height (vowel height) than in the front-back dimension for vowels. In this case, the preference for changes in vowel height is regarded as evidence that variation in mandibular activity is the primary source of phonetic variation. According to Davis and MacNeilage (1995, p. 1208), this variation consists of changes in the amplitude of mandibular oscillation.

The inference of specific patterns of mandibular oscillation from vowel height is normally based on the assumption that vowel height is determined by the level of the first formant frequency ($F_1$), and that the level of $F_1$ is itself determined by the amplitude of mandibular oscillation. It would appear, however, that the level of $F_1$ is less closely linked to vowel height than was once assumed and that this is particularly true of infant vowels. This can be seen most clearly in the phenomenon of overlapping vowel-spaces in the $F_1$-dimension (Lieberman 1980). Furthermore, studies of the perception of adult vowels show that a variety of acoustic features are used to differentiate between vowels according to vowel height, including those relating to fundamental frequency, vowel inherent spectral change, and nasalization (Di Benedetto 1989; 1994; Kingston & Macmillan 1995). The preliminary results of a study by this commentator show that in the case of infant vowels, as well, perceived vowel height may be determined by a combination of acoustic features and not simply by the level of $F_1$ attained (Peters, in preparation).

Even if the perceived vowel height of infant vowels is to be determined solely on the basis of the level of $F_1$, changes in $F_1$ are not necessarily caused by changes in the amplitude of mandibular oscillation. Fixed-mandible experiments carried out on adults (Lindblom et al. 1979) have shown that variations in $F_1$ can be achieved even without any movement of the mandible. Moreover, in the case of adults, mandibular movements are regarded as having only a supportive role in the production of speech. Such movements are normally coordinated with movements of the tongue and lips, whereby the contribution of the respective articulatory organs may vary from person to person. Even in the first year of a child's life, there is evidence of basic coordination of articulatory movements. For example, the emergence of rounded vowels suggests that lip movements are being coordinated with movements of the mandible or tongue. In addition, the fact that utterances, which, according to hypothesis (5) are not preferred, did occur in MacNeilage's data, suggests that the infant may nonetheless possess the ability to combine movement of the tongue with movements of the mandible, even if this is not the infant's preferred means of production.

Even considering the other preferences outlined in section 5.2 regarding the segmental composition of babbling utterances, it is not necessarily the case that movements of the mandible are primarily responsible for phonetic variation. Both the preference for a combination of vowels and consonants produced at similar places of articulation (hypotheses 1–2) and the preference for intersyllabic changes in the manner of articulation over place of articulation for consonants (hypothesis 4) could be expected, even if the infant were able to coordinate tongue and mandibular movements to vary the height of the tongue. Furthermore, as far as hypothesis (3) and the assumption of "pure frames" is concerned, it is debatable whether the tensing of the lips that is necessary to produce labial stops such as [b] can be realized solely by means of the upward pressure of the mandible.

In conclusion, the assumption that variations in mandibular movement are primarily responsible for phonetic variation in babbling utterances may well be correct. However, transcription-based evidence alone does not enable us to determine the actual extent of the role played by mandibular movement in the production of utterances by the child. It is therefore necessary to re-examine the argument put forward in section 5.2 on the basis of acoustic and articulatory data.

## What happened to *Homo habilis*? (Language and mirror neurons)

Giacomo Rizzolatti

*Instituto di Fisiologia Umana, Università di Parma, 43100 Parma, Italy.*
**fisioum@symbolic.pr.it**

**Abstract:** The evolutionary continuity between the prespeech functions of premotor cortex and its new linguistic functions, the main thesis of MacNeilage's target article, is confirmed by the recent discovery of "mirror" neurons in monkeys and a corresponding action-observation/action-execution matching system in humans. Physiological data (and other considerations) appear to indicate, however, that brachiomanual gestures played a greater role in language evolution than MacNeilage would like to admit.



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

I like this target article very much. Having long been involved in the study of the ventral premotor area F5, the probable monkey homologue of Broca's area, I enjoyed seeing the evolutionary continuity between the prespeech functions of premotor cortex and its new linguistic functions being spelled out so persuasively.

Physiological evidence, from "mirror" neurons in particular, clearly supports this continuity. Mirror neurons are a set of F5 neurons that discharge *both* when the monkey performs an action and when it observes another individual performing it (Gallese et al. 1996; Rizzolatti et al. 1996a). Mirror neurons appear to form a system that allows individuals to recognize motor actions made by others by matching them with an internal motor copy (Carey et al. 1997; Rizzolatti et al. 1996a). Transcranial magnetic stimulation (Fadiga et al. 1995) and PET experiments (Decety et al. 1997; Grafton et al. 1996; Rizzolatti et al. 1996b) indicate that a similar mechanism is also present in humans and involves, among other areas, Broca's area.

There is an important point, however, on which I do not agree with MacNeilage: his strong bias against "signed" language and its importance for language evolution (sect. 7.6). Although the main theme of the target article is speech production, and "signed" language might therefore seem to be only a side issue, I do not think it is. The reason will become clear after considering MacNeilage's position and mine on why *Homo habilis* had "speech" areas (Falk 1983; Holloway 1985; Tobias 1987).

According to MacNeilage the presence of a Broca-like area in *Homo habilis* testifies that "the main change from other primate vocalization to human speech has come in the articulatory system" (sect. 6.1). If I interpret his thinking correctly, at a certain evolutionary stage, the cortical medial system, which mediates primate calls, became insufficient for interindividual communication. It therefore "primed" the monkey premotor homologue of Broca's area to assume this role. This area, because of its preexistent connections with the primary motor cortex and subcortical centers, had the anatomical substrate for achieving a refined control of the vocal tract. Thus, a ventral premotor area located on the lateral cortical surface and originally used for nonlinguistic purposes became the motor speech area in *Homo habilis.*

My view is different: the growth in *Homo habilis* of the frontal (Tobias 1987) and temporo-parietal cortical regions (Falk 1983; Holloway 1981) was not caused by an evolutionary pressure for speech production (which in *Homo habilis,* at least judging from its very primitive vocal tract, was very limited; Lieberman 1984), but a consequence of the development of the action-observation/action-execution matching system (mirror system). This system (areas with mirror neurons and areas related to them) is located in monkeys and humans in those cortical regions that developed in *Homo habilis.* A possible hypothesis is, therefore, that these regions evolved because of a greater need to recognize actions in hominids than in nonhuman primates. This need was in turn determined by hominids rapidly growing motor repertoire.

The basic function of the mirror system is that of recognizing actions by others (Carey et al. 1997; Rizzolatti et al. 1996a). Later in evolution, possibly in the transition from *Homo habilis* to *Homo sapiens,* the mirror system started being used for intentional communication (for a discussion of this point, see Rizzolatti & Arbib 1998). Which types of movements were used for this purpose, orofacial or brachiomanual movements? The answer is most likely both. Although the arguments that MacNeilage advances for including orofacial movements in the evolutionary route to speech are convincing, they do not rule out brachiomanual gestures. Against a purely orofacial hypothesis of speech evolution, I see two objections. The first is that in orofacial communication, the exchange of communication is essentially limited to two individuals. The possibility of introducing a third element is very limited. The second objection lies in the fact that the combinatorial properties, which represent among the most important properties of speech, are virtually absent in orofacial communication. In contrast, they are inherent to the brachiomanual system (see Arbib & Rizzolatti 1997; Corballis 1992), in both its intransitive and transitive use (actions directed toward objects).

This scenario, although very sketchy, explains the fact that human centers for language are located on the lateral cortical surface (perisylvian region plus Broca's region) and not medially, as are the call centers in primates. It explains the paradox of *Homo habilis* who, although endowed with an anatomically well developed Broca's area, had a vocal tract that could hardly subserve an elaborated motor control. In addition, the presence of different fields in F5 (and in Broca's area) for orofacial, brachiomanual, and orolaryngeal movements strongly supports the contention that interindividual communication did not evolve from a single motor modality, but resulted from an interplay of facial gestures, brachiomanual gestures, and, finally, sound gestures. (As far as the latter are concerned, it is worth noting the close similarity between the mirror mechanism and that proposed by Liberman & Mattingly [1985] for speech perception.)

In conclusion, I fully agree with MacNeilage that sound communication derived from an evolutionary "tinkering" of the preexistent structures originally developed for ingestive purposes. His frame/content theory is a beautiful example of how a function that, like speech production, appears to be somehow magically unique can be traced back to older and much lower functions. Yet I also think that the pathway from these functions to the new ones was much more tortuous than it appears from the target article. The human capacity to communicate developed from progressive, global evolution of many mirror systems, not just the orofacial one.

## Ingestive and vocal mechanisms in birds: A parallel?

Jim Scanlan and Lesley Rogers

*Division of Neuroscience and Animal Behaviour, School of Biological Sciences, University of New England, Armidale, NSW 2351, Australia.*
**jscanla2@metz.une.edu.au**

**Abstract:** Parrots prepare for vocalization by a ventro-caudal retraction of the larynx. This laryngeal movement, which "frames" vocal sequences, is similar to a movement used by pigeons as a preparation for suction drinking. The air-pressure events involved in such movements can trigger either suction drinking or vocalization. This suggests a possible evolutionary link between these ingestive and vocal mechanisms.

We would like to draw attention to ingestive mechanisms in birds that, when related to vocal mechanisms, may support a more general application of the "frame/content" theory. Our observations concern movements of the larynx. Although studies of laryngeal movements have concentrated on feeding functions, we have observed comparable movements during vocalization.

X-ray films showing movements of the larynx in an African grey parrot (*Psittacus erithacus;* film courtesy of Professor R.-G. Busnel) and a cross-bred Amazon parrot (*Amazona spp;* film courtesy of Professor G. Du-Boulay) reveal two main kinds of laryngeal movements associated with vocalization: preparatory movements and synchronic movement. Preparatory (or prevocalizing) movements are gross movements of the lingual apparatus that transport the larynx to a vocalizing position; synchronic movements occur during vocalization itself. By far the most common preparatory movement is a ventro-caudal retraction of the larynx. This movement, which is conducted in silence, is always followed by a strong and sudden onset of sound and is similar to the preparatory movement observed in the crowing of roosters by White (1968a) and White and Chubb (1968). In its exclusive connection with the onset of sound, and not with sound modulation within a continuous vocal sequence, this movement may have affinities with the "framing" movements of MacNeilage's theory.

The published studies of laryngeal movements during ingestion (Homberger 1980; White 1968b; Zweers 1982a; 1982b; Zweers et


https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

al. 1981) reveal a comparable distinction between those that prepare the oral-pharyngeal cavity for taking in liquids or solids (preparatory movements) and those that actively transport food or drink toward the esophagus (synchronic movements). Synchronic movements include rostro-caudal activity, in which the larynx and its papillae act as a "rake" in propelling liquids and solids along the pharyngeal roof. An important preparatory movement, on the other hand, is that observed by Zweers (1982a) during suction drinking in pigeons. This is a caudal movement of the larynx just before the entry of liquid into the pigeon's mouth. There is a strong visual similarity between this preparatory laryngeal movement in pigeons' drinking and the preparatory movement we have observed in the vocalization of parrots. This suggests a possible connection between an ingestive mechanism and a vocal "framing" mechanism, as in MacNeilage's theory.

There could be functional as well as visual similarities between the ingestive and vocalizing movements. Zweers (1982a) explained the preparatory suction-drinking movement, with its consequent caudal and ventral extension of the buccal air space, as functioning to reduce pressure within the buccal cavity, thus forcing liquid into the mouth under atmospheric pressure. Because the glottis is necessarily closed during drinking, its caudal movement in preparation for water-suction must be associated with glottis closure. With the glottis closed, a reduction of supraglottal pressure could produce a significant pressure difference in addition to the intra/extra buccal difference described by Zweers. This is a pressure difference across the laryngeal valve: a pressure difference that – considered in the context of vocalization rather than of drinking – would facilitate a sudden explosive release of subglottal air and thus initiate vocalization.

Current understanding of avian drinking mechanisms does not link suction drinking in pigeons directly with the ladling method used by the Psittacinae (Homberger 1980), and the phylogenetic relationship of the two drinking methods is unclear. However, if there is an evolutionary link between suction drinking and the preparatory laryngeal movement in parrots, the parallel with MacNeilage's theory would be further strengthened. It could represent an addition to the list of intriguing analogies between humans and birds in both the phylogeny and the ontogeny of vocalization.

This avian analogy raises the possibility of a more widespread use of ingestion-related mechanisms in vertebrate vocalization. For example, although the target article argues that the chewing-articulation relationship is a uniquely human adaptation, Hauser et al. (1993) have observed systematic movements of the lips, jaw, and teeth during vocalization in rhesus monkeys that can be correlated with consistent changes in formant frequencies (but not in fundamental frequency). This indicates a complex interaction of phonation and articulation comparable to that in speech, and raises the question of a possible phylogenetic relationship between the two vocal systems.

In mammals generally, interaction between ingestive mechanisms and respiratory/vocal functions appears to have been a major influence on the evolution of vocal behaviour: the larynx developed its phonatory function after evolving as a valve to protect the airway during swallowing. A similar interaction in the subsequent (and complementary) development of vocal articulation, such as that proposed in humans by MacNeilage, could thus have phylogenetic ramifications, at least within the primates. Our knowledge of primate vocalizations is still not sufficient to preclude such a possibility.

## Recent evidence of the involvement of lateral frontal cortex in primate cyclic ingestive movements

Barry J. Sessle

*Faculty of Dentistry, University of Toronto, Toronto, Ontario, Canada M5G 1G6.* **barry.sessle@utoronto.ca**

**Abstract:** This commentary focusses on MacNeilage's arguments and evidence that the development of cerebral cortical controls over cyclic ingestive movements has provided substrates for the evolution of speech production. It outlines evidence from experimental approaches using cortical stimulation, inactivation, and single neuron recording in primates that lateral frontal cortical regions are indeed crucial for the generation and guidance of cyclic orofacial movements.

MacNeilage provides some provocative but well-argued proposals on the evolution of speech production. This commentary focuses on the proposition (sects. 4 and 6) that this evolution is associated with the development of cerebral cortical controls over the same central pattern generators (CPGs) that contribute to cyclic ingestive behaviors. MacNeilage refers to certain predictions stemming from the testing of his theory, including the representation of ingestive cyclic movements in different cortical regions (sect. 7.1). Although it is presently unclear to what extent the medial cortical regions mentioned by MacNeilage are involved in these movements, it should be noted that evidence already exists for extensive representations of cyclic orofacial movements in certain parts of the lateral frontal cortex. This cortical region, and in particular the primary motor cortex (MI), has long been known to be involved in the initiation and control of movements. Specifically, in the case of orofacial movements, cortical surface electrical stimulation or intracortical microstimulation (ICMS) of face MI in several mammalian species evokes twitch-like movements of the facial, jaw, tongue, and laryngeal/pharyngeal muscles (see Luschei & Goldberg 1981; Martin & Sessle 1993; Martin et al. 1997).

Mapping with ICMS of monkey face MI reveals extensive facial, tongue, and (to a lesser extent) jaw-opening motor representation; ICMS-evoked jaw-closing movements are extremely sparse, and neither lesions nor inactivation of face MI have much of an effect on biting behavior, although they severely disrupt other trained orofacial movements (see Luschei & Goldberg 1981; Martin & Sessle 1993). Nonetheless, cortically generated, rhythmic jaw-closing and jaw-opening movements do occur. Repetitive electrical stimulation of parts of the anterolateral frontal and lateral pericentral cortex evokes cyclic jaw movements in a number of species, including humans and monkeys (see Huang et al. 1989; Luschei & Goldberg 1981; Martin & Sessle 1993; Martin et al. 1997). These movements resemble mastication and are frequently accompanied by licking, sucking, or swallowing. For example, ICMS evokes rhythmical jaw movements reflecting several different masticatory patterns, as well as swallowing, from four discrete cortical regions: the face MI, the primary face somatosensory cortex (SI), the classical "cortical masticatory area" (CMA) lateral to MI, and a deep CMA located on the inferior face of the frontal operculum. Also of interest to MacNeilage's discussions (sects. 4.3 and 5.2) of ontogenetic and phylogenetic features bearing on speech development is the finding (Iriki et al. 1988) that stimulation of the so-called cortical sucking area (CSA), in the area rostral to CMA, induces rhythmical sucking-like movements in neonatal guinea pigs, whereas the conversion from sucking to mastication during the maturation of feeding may be associated with a shift from CSA to CMA in the cortical projection to brainstem regions involved in cyclic ingestive behaviors.

This documentation of the extensive cortical representation of primate ingestive cyclic movements is supported by findings that trauma, surgical ablation, or reversible inactivation by cold block of those regions from which cyclic ingestive behaviors can be elicited, severely disrupts mastication and swallowing (see Lin et al., in press; Luschei & Goldberg 1981; Martin & Sessle 1993).



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

Furthermore, the discharge patterns of single neurons recorded in the awake monkey's lateral pericentral cortex, including Brodmann's areas 4 and 6 and the CMA, are related to a variety of orofacial movements. These include trained biting or tongue protrusion behaviors and of particular relevance to MacNeilage's theory and predictions are findings that the discharges of many face MI, SI, and CMA neurons may be related to cyclic ingestive movements associated with licking, mastication, and/or swallowing (e.g., Luschei & Goldberg 1981; Martin et al. 1997; Murray & Sessle 1992b).

It is also notable that face MI and CMA neurons may receive extensive orofacial afferent inputs, with close spatial matching of inputs and outputs especially in face MI. Recent findings indicating that reversible cold block of the monkey's face SI disrupts both trained tongue motor behavior and cyclic ingestive movements associated with chewing and swallowing (e.g., Lin et al. 1997) underscore the importance of these somatosensory inputs and motor effects in cortical mechanisms contributing to the guidance, control, and learning of orofacial movements (Huang et al. 1989; Murray & Sessle 1992b). They support MacNeilage's view (sect. 7.5) of the significance of input-output linkages in the evolution of the control mechanisms underlying speech. Such linkages also bear on his earlier point (sect. 5.2) about human babbling and the development of frame and content. This development occurs at a time when the first teeth erupt (6–7 months of age). The exquisite sensorimotor control properties that dental mechanoreceptors confer on the masticatory system through their brainstem projections and cortical input-output linkages might provide an additional structural and functional framework for rhythmic orofacial movements to develop and provide guidance for the acquisition of articulatory as well as masticatory skills.

Considerable evidence is emerging, therefore, that the primate frontal cortex, including face MI, has evolved to provide important neural mechanisms that serve to initiate and guide not only trained or learned motor behavior, but also cyclic ingestive movements. These findings are consistent with MacNeilage's view that higher control systems have evolved to provide additional control mechanisms above and beyond the brainstem CPGs. The frequently enunciated view that MI plays a major role in controlling an operantly conditioned movement such as a biting task, but defers to the CPGs and plays only a minor role in the control of cyclic movements involving the same muscles, may need reassessment in light of the above-mentioned findings pointing to a significant role of primate MI as well as CMA in cyclic ingestive movements.

ACKNOWLEDGMENTS

Studies by the author were supported by Canadian M.R.C. grant MT-4918.

## An evolutionary model for the learning of language

Jechil S. Sieratzki[a] and Bencie Woll[b]

[a]*Dana Children's Hospital, Department of Pediatrics, Sackler School of Medicine, Tel Aviv University, Ramat Aviv 69978, Israel;* [b]*Department of Clinical Communication Studies, City University, London EC1V OHB, England*
**bguest@econ.tau.ac.il; b.woll@city.ac.uk**

**Abstract:** This commentary deals with the relation between human language and nonverbal signals used by nonhuman primates. It suggests that human language could have developed through the interaction of procedural learning with a preexisting system for socio-affective communication. The introduction of "content" into existing "frames" requires a neurobiologically plausible learning mechanism.

Nonverbal aspects of speech have an important place in human language (Ross 1993). Speech melody (prosody) has a crucial signal function beyond the bare informative content of words. Prosodic modulation is particularly important in early mother-infant interaction, creating a proto-language that is remarkably similar in all cultures (Sieratzki & Woll 1996). Even deaf mothers initially vocalise to their deaf infants, although neither can hear the sound (Woll & Kyle 1989).

It is reasonable to assume that the limited communicative repertoire of nonhuman primates is controlled by both hemispheres, in comparison with the cerebral asymmetry of human language (Hellige 1993). In humans, the nondominant hemisphere controls nonverbal socio-affective aspects of speech ("frames"; Ross 1993), whereas the dominant hemisphere has become specialised for creative verbal content and grammatical organisation in the place of fixed signals ("frames" plus "content"). During hominid evolution, not only a more complex phonation system but also a new analytical and combinatorial learning capacity emerged.

We hypothesise that human language originated from the interaction of existing forms of associative-procedural learning with a preexisting system for the production of socio-affective communication. Procedural learning occurred early in evolution and develops early in human infants (Tulving 1995). Children and young animals are exhilarated by sensorimotor, action-related exploration of their environment. Human toddlers develop an early interest in examining the different shapes of objects. In contrast with nonhuman animals, these first steps lead to a rapidly maturing ability to categorise spatial objects as structural, physical, perceptual entities (Tulving 1995).

This ability has not been observed in preverbal humans (Tulving 1995) and appears to be functionally relevant for the development of language. For example, research on motor-impaired children with spinal muscular atrophy, who show a striking precocity in over-regularisations, suggests that children learn grammatical forms in procedural steps, like objects of spatial learning (Sieratzki & Woll 1998). More generally, language development exhibits the typical characteristics of the learning procedures of early childhood – active exploration, repetitive exercise, internal correction and regularisation, and the generation of rule-governed output mechanisms – combined with reciprocal socio-affective interaction.

Learning has obvious innate constraints. To advance to language, previously existing hominid gestures and vocalisations must have become detached from a framework of preprogrammed reflexive responses and evolved into objects of internal categorisation with linguistic meaning. Reciprocal exchange between able individuals enhanced exponentially the advance of this process. A developing ability became incorporated into expanding neocortical areas: a learning mode developed into a neural module.

It is unclear under what circumstances progress toward the structural organisation of communication began: Did an emerging ability for spatial categorisation open the gates for symbolic manual gestures accompanied by vocalisations, which then evolved into noniconic speech (for a more detailed explanation see the commentary by Woll & Sieratzki)? What does seem certain, however, is that once the first step was taken, the mastery of language was not related to mankind's mastery of the physical environment. Languages found in stone-age cultures are most complex and complete, but dissociated from the level of technological development.

We must therefore regard language as both tool and target of its own development: a perpetual evolutionary dynamic, which, like an environmental condition shaping a biological trait, expanded its own neural basis, and with it learning, memory, and the capacity of the human mind. Language is the bridge early hominids built to become Homo Sapiens.

To compare this development to evolutionary tinkering underestimates the intrinsic dynamic of this process; on the other hand, to imply an instantaneous onset resulting from a single mutation moves the gift of language into the inexplicable realm of divine creation, away from plausible neurobiological theory.



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

## Brain circuits ancient and modern

Stephen F. Walker

*Centre for Physical and Life Sciences, Birkbeck College, London WC1E 7HX, England.* **s.walker@psychology.bbk.ac.uk**
**www.psyc.bbk.ac.uk./staff/sfw.html**

**Abstract:** I support the application of the "evolution as tinkering" idea to vocalization and emphasize that some of the subcortical parts of the brain circuits used for speech organs retain features common to nonprimate mammals, and in some cases to lower vertebrates, pointing up the importance of cortical evolution as suggested by MacNeilage.

No one can disagree that the articulatory organs (tongue, jaws, lips, larynx) are also used in eating. The more peripheral organization of neuromuscular control of these organs (e.g., the functions of the cranial nerves; Hamdy et al. 1997; Lazar et al. 1992) is relatively conservative in vertebrate evolution; hence some aspects of the neural control of speech can be related not only to brain mechanisms common to humans and nonhuman primates, but also to nonprimate and even nonmamalian cross-species comparisons.

There is, for example, a paradox in the fact that a comparison of musculotopic organization of the hypoglossal nucleus in the grass-frog (Sokoloff 1991) and the macaque monkey suggests a high degree of evolutionary conservatism in this part of the control of tongue movement (Sokoloff & Deacon 1992), whereas the use of the tongue in human speech is sufficiently distinctive to have given rise to the theoretical problem for which MacNeilage offers a solution.

It would be consistent with his theme to try to resolve this paradox partly by appealing to a process of corticalization of vocalization, with language evolution seen as the addition of voluntary and learned cortical control of vocal communication to the fixed subcortical or limbic cortical elicitation of innate acoustic signals (Walker 1994). The hypothesis would have to be that voluntary and learned cortical control of eating and some forms of head and limb movement is characteristic of mammals, but that these aspects of neocortical potential only became fully applied to vocalization at some point during specifically human evolution. Modern brain imaging techniques should allow hypotheses about human cortical and subcortical function to be tested (e.g., Urban et al. 1996).

In cats (Zhang et al. 1995) as well as monkeys (Jürgens 1994), much of the muscular coordination required for vocalization is subcortical, involving circuits through the midbrain periaqueductal gray (PAG) and brain-stem nuclei. Davis et al. (1996) have recently proposed that the PAG can be classed as a universal brain site for mammalian voice production, generating respiratory and laryngeal motor patterns for both emotional and involuntary sounds and for human speech and song. Jürgens & Zwirner (1996), however, suggest that the PAG is a relay station for limbic or emotional vocal expression, but not part of the route for neocortical output from facial motor cortex.

There is evidence that control of some kind of oral activity was a very early feature of the functions of mammalian cortex, because multiple somatosensory cortical representation, which includes the orofacial regions, has been demonstrated in monotremes (Krubitzer et al. 1995) and marsupials (Beck et al. 1996). Dual pathways of output from orofacial motor cortex are observed in guinea pigs (Enomoto et al. 1995), and rats have a specialized "jaw, lips, and tongue" region of motor cortex (Ebrahimi et al. 1992). Macaque monkeys can be trained in a variety of tongue-protrusion tasks, and the participation of individual neurons in both sensory and motor cortex in these tasks can be monitored (Lin et al. 1994; Murray & Sessle 1992a). This is consistent with the notion that there is voluntary cortical control of the tongue and lips for ingestive movements in nonhuman primates, but not fine cortical control of the coordination of the articulators in vocalization (Hayes & Nissen 1971). Tongue movements in the above studies appear to be localized in motor cortex separately from jaw movements (Murray & Sessle 1992). It supports MacNeilage's thesis that, within cingulate cortex, recording of single unit activity suggests that some neurons relate to both jaw-opening and vocalization; there are also neurons that specialize in just one of these activities (West & Larson 1995).

How the flexible control of ingestive tongue and jaw movements in primates with fixed patterns of vocalization developed into human speech remains a puzzle, and the suggestion that lip-smacks and teeth chattering provide intermediaries is at least as well supported as the alternative of hand and arm gestures (Wilkins & Wakefield 1995).

## Echo phonology: Signs of a link between gesture and speech

Bencie Woll[a] and Jechil S. Sieratzki[b]

[a]Department of *Clinical Communication Studies, City University, London, EC1V OHB, England;* [b]*Department of Human Communication Science, University College London, London WC1, England;* **b.woll@city.ac.uk; bguestecon.tau.ac.il**

**Abstract:** This commentary supports MacNeilage's dismissal of an evolutionary development from sign language to spoken language but presents evidence of a feature in sign language (echo phonology) that links iconic signs to abstract vocal syllables. These data provide an insight into possible mechanism by which iconic manual gestures accompanied by vocalisation could have provided a route for the evolution of spoken language with its characteristically arbitrary form–meaning relationship.

The recent resurgence of interest in the origin and evolution of language has led to the suggestion that sign languages might represent an earlier stage of human language than spoken languages. Researchers such as Armstrong et al. (1994) have argued that the transition from gesture to sign language preceded the development of spoken language. In contrast, Sieratzki and Woll (1996) and Woll (1996) have argued that although neural plasticity creates the equipotentiality for a child to develop either signed or spoken language, evidence supports the view that (1) for modern Homo Sapiens, spoken language has primacy over sign language, (2) human language developed first in the auditory modality, and (3) sign language as a linguistic system developed after spoken language, an argument in accord with MacNeilage (1987a).

MacNeilage himself, in the target article, rejects his earlier model (1987a) in which manual activity precedes and develops into vocal activity, because of the difficulty of postulating a mechanism for such a transfer to take place. Another reason underlying the rejection of sign language as a precursor of spoken language has been the difficulty of seeing how the largely iconically motivated signs of sign language could have been transformed into the largely arbitrarily motivated words of spoken language. Nevertheless, such a mechanism can be observed in sign language, in what we have called "echo phonology," in a group of oral components found in British Sign Language (BSL), and other sign languages (Lawson 1983; Pimiäa 1990; Schermer 1990; Vogt-Svendsen 1983; Woll 1993).

The term echo phonology is used because the oral movement components found in this group of signs mirror or echo the manual movements. For example, in BSL, abrupt separating of the hands is accompanied by the oral syllable [pa]. We must assume for these examples, as they are not derived from spoken language, that the hands "drive" the mouth, and not the other way around, as in gestures accompanying speech.

These elements are obligatory in citation forms of certain lexical signs, and are neither derived from spoken words nor visually motivated. All examples require the exhalation or inhalation of breath, usually with a change in mouth configuration during the articulation of the sign (rather than static mouth arrangements



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

such as "tongue protrusion," which are also found in sign languages, but associated with adverbials). Echo phonology has a structure that can serve to support theories about syllables in the manual component of sign languages and can also provide clues about the possible origins of spoken language phonology.

Syllables identified thus far in BSL include [pa-phonetic] (occurring with separating hands), [ʃʃ] (occurring with finger or hand oscillation), [ʌp] (occurring with movements where the hand closes and approaches the body), [ʌm] (occurring with movements where the active hand contacts the passive hand), and [θʊp] (occurring with closing hands). Where the hands are temporarily occupied, these syllables can occur (and are understood) on their own.

The oral activities in echo phonology are not themselves iconic. It is impossible to reconstruct from the echo syllable [ʃʃ], occurring with the signs *exist, not-yet,* and *substantial,* any common visual motivation, although the manual activities can be interpreted as visually representing the marking of a small area in space, a dismissive side-to-side shaking of the hands, and the description of something of large size, respectively. The only feature common to all three signs is a small oscillating wrist or finger movement, which is echoed in the oral action.

While still wishing to argue that sign language did not precede spoken language, echo phonology data provide an insight into a possible mechanism by which manual gestures, accompanied by vocalisation, could have provided a route for the development of spoken language.

# Author's Response

## The frame/content view of speech: What survives, what emerges

Peter F. MacNeilage

*Department of Psychology, University of Texas at Austin, Austin, TX 78712;*
**macneilage@mail.utexas.edu**

**Abstract:** There was little disagreement among commentators about whether speech production involves a frame/content mode of organization, but there was some disagreement with the contention that frames evolved from ingestive cyclicities and were mediated via a medial "intrinsic" system.

### R1. Introduction

The theory presented in the target article had two main tenets: (1) Speech has evolved a frame/content mode of organization. (2) The frame component is more associated with a medial "intrinsic" neural system whereas the content component is more associated with a lateral "extrinsic" system. Nobody took explicit issue with the appropriateness of the frame/content dichotomy for modern adult speech production. Nor, with one exception, did anyone take issue with the view that frames precede content in ontogeny. However, there was some dissent from the claim that frames had an ingestive origin. A few commentators were dissatisfied with the allocation of frames to the medial system, and there was even some dissent as to whether the cortical premotor part of the lateral system was of any importance. Surprisingly little attention was given to the possibility of an important role for visual gestural communi-cation in language evolution. However, a number of commentators provided a perspective on the frame/content view by considering vocal-auditory communication in other taxa. These topics will form the main framework of this Response, followed by comments on some remaining issues.

### R2. The frame/content mode in modern hominids

**R2.1. The F/C mode in adult speech.** The lack of an explicit objection to the frame/content characterization of modern speech production would seem to be a very important outcome of the target article. The credit for the survival of this dichotomy goes primarily to the branch of psycholinguistics concerned with speech errors at the phonological level. This outcome gives support to my contention (in sect. 7.2) that "no theory of . . . the organization of speech . . . that does not include the dual components metaphorically labelled frame and content . . . is a viable one."

**Levelt & Schiller** provide recent evidence regarding the representation of frame information in the mental lexicon, and in the process list a veritable Who's Who of the pioneers in speech error studies (I would add MacKay 1970). Their contention is that in languages such as Dutch and English, which have a dominant stress pattern, lexical items, in addition to having stored representation of number of syllables, have only stored metrical frames when they deviate from the default stress patterns. See Levelt et al. (in press) for a comprehensive view of lexical access with implications for brain imaging studies.

**R2.2. Ontogeny of the F/C mode.** The only objection to the frame/content perspective on modern speech production was directed at speech acquisition. In this realm, without denying **Menn**'s contention that other articulators are also involved, that there are individual differences in speech acquisition patterns, and that there are language-specific influences even on babbling content, Davis and I contend that the syllable-related frame provided by oscillation of the mandible is the primary source of motor variance in babbling and early speech (MacNeilage & Davis 1990a; 1993). More recent evidence for this view will be given later. Menn dismisses our perspective by concluding that "Articulatory phonology . . . though limited to a motor focus, can capture the emergent character of the syllable in a way that other approaches cannot."

There is no substance to this claim, though it is unfortunate that **Menn** did not lay out the supposed basis for it. The unit of articulatory phonology is the "gesture" (Browman & Goldstein 1986). The definition of the gesture is in terms of the formation *and release* of a constriction (emphasis mine), but this definition has not yet been shown to reflect articulatory reality in that these two distinct phases of articulatory movements have not yet been shown to have functional unity. The approach is basically ad hoc in that it begins with observable constrictions and releases in adult speech and then gives them conceptual status by means of reification. To date, nothing analogous to a separable frame component has been identified in articulatory phonology, even though the mandible has been shown to be an independent variable at the adult motor level, in that mandibular depression varies as a function of contrastive stress on a syllable (Erickson et al. 1994). In extending the ad hoc mode of analysis of articulatory phonology to infants, it


https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

would seem that an infant utterance such as [baba] would need to be analyzed primarily as two iterations of a labial gesture followed by a low central tongue gesture, although there is no evidence that the consonantal and vocalic components of this utterance are functionally independent or that the lips or tongue are even activated for these events. **Peters** makes the point that our perception-based claims regarding movements are hypothetical. However, **Munhall & Jones** report that, as we have suggested, there is an absence of independent lip activity in the labial consonants of one babbling subject. There is a further problem. From the ad hoc standpoint, the labial consonant and low central vowel gestures in [bababa] and the coronal and low front gestures in [daedaedae] must be considered to be different. But according to the "frames, then content perspective for speech acquisition, there may be no difference in active movement control during the production of these two utterances. The difference may only result from the adoption of a tongue front position for the latter before it even begins.

In summary, contra **Menn,** the frame/content theory cannot be dismissed without comment as a perspective on speech acquisition and replaced with one which has neither given serious attention to the syllable at the premotor or the motor levels, nor provided a treatment of babbling and early speech that is not simply ad hoc. In fact, as articulatory phonology applies only to the content component of speech, the approach is even less able to characterize babbling and infant speech (where frame influences seem to be dominant) than it can adult speech.

Work done since the target article was submitted provides additional evidence for the view that babbling and early speech are best characterized by frame dominance. The consonant-vowel co-occurrence constraints found in babbling (coronal consonants with front vowels, dorsal consonants with back vowels, and labial consonants with central vowels) have also been found to be present for the most part in the so-called 50-word stage (12–18 months) of early speech of the subjects already studied by Davis and MacNeilage (1995) during babbling (MacNeilage et al. 1997). We have also found the same intersyllabic vowel and consonant variegation preferences as in babbling during this stage – consonants vary primarily in manner and vowels primarily in height.

The fact that we have also found these trends in data from first word production in a group of subjects that we did not phonetically transcribe (MacNeilage at al. 1997) should partially allay **McCune**'s concern that we are the only group who have consistently reported these trends to date. In addition, an acoustic study of babbled syllables shows that rather than transcribing in such a way as to favor our hypotheses, we sometimes tend to do the opposite. An analysis of transcribed vowels that did not fit the hypothesis (front vowels with labials and central vowels with coronals) showed that even these vowels tended acoustically towards the predicted trends. Vowels transcribed as central in coronal environments had higher second formants (suggestive of a more front tongue position) than those transcribed as central in labial environments. Conversely, vowels transcribed as front vowels in labial environments had lower second formants than similarly transcribed vowels in coronal environments (Davis & MacNeilage 1995). We have virtually no doubt that when databases as large as the ones we have used are studied in individual subjects, most infants will show these indications of frame dominance.

Although evidence for frame dominance at the motor level in babbling and early speech is extremely strong, this does not necessarily mean that motor factors are more important than perceptual factors in speech acquisition. **Menn** mistakenly attributes this view to me, not noticing that I concluded (sect. 6.4) that "from a perceptual-motor perspective the main change in vocal organization from other primates to humans may be evolution in the LPS of a capacity to *learn* speech." In a mimetic process, as is required for speech acquisition, perception is of course crucial, but it is meaningless without production. At a general level, production and perception must be integral to each other in speech acquisition in modern hominids, and if one is to be considered more important than the other it must be for some specific aspect of the overall process. Menn emphasizes the well-known fact that a normal pattern of babbling does not occur in infants with severe hearing loss. However, my colleagues and I have recently observed that on the infrequent occasions when one such infant produces CV alternations, labial consonants co-occur with central vowels and coronal consonants with front vowels, and this indication of frame dominance is presumably not based on a perceptual analysis of adult CV relationships (McCaffrey et al. 1997). Thus, relatively normal perceptual abilities are crucial to normal babbling development, but at least some aspects of the particular babbling patterns produced may be primarily determined by motor constraints.

**Bloom** adds information relevant to the time course of the ontogeny of frames. She believes that frame-like production may begin as early as three months of age in the form of what she calls "syllabics," a class of utterance with phonation and an open mouth, in which the mouth is "frequently moving" (Bloom 1989, p. 246). The movements involved here may require more scrutiny before being considered precursors to the frames of babbling. Bloom also points out that there have been informal observations of mouth open-close alternations without phonation, as early as three months of age.

## R3. Evolution of the F/C mode

The lack of any plausible reason for rejecting the F/C perspective for either modern adult speech or speech ontogeny enhances the legitimacy of the enquiry into its origins. Here the sailing it not so smooth. Of all the commentators, **Lindblom** is most in sympathy with the proposal that the frame provided by mandibular oscillation could have been exapted from ingestive processes and used for communicative purposes. Like me, he is impressed by the unanimous choice of the frame in the world's languages in the presence of so many other articulatory possibilities, and its consequent importance from an explanatory perspective. However, he stresses the fact that simple exaptation is unlikely to be enough, rejecting the view common in linguistics that the advent of some new constraint which was arbitrary relative to function would have been sufficient for speech evolution. Further tinkering into a form most appropriate for new use is also needed, and in the case of speech this must have included adaptation to the fact that speech requires less force from the mandible than ingestive processes.



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

**Lindblom**'s basic thesis has always been that the structure of speech is the result of a continuous tug-of-war between patterns that are compatible with basic motor capabilities, and the demand for a family of perceptually distinctive patterns (e.g., Lindblom 1992). Consequently, in his opinion, perceptual constraints must have played a major role in the tinkering with any basic early motor patterns to produce present day speech structures. Although the target article concentrated on motor factors in speech evolution, this was not done with the intention of minimizing the importance of perceptual factors in the evolutionary process. I have argued that sounds acquired later by infants were probably introduced into languages at a stage beyond evolution of the earliest sound patterns under the impetus of developing increasing perceptual distinctiveness (MacNeilage 1994). Although the frame may have initially been selected for communication primarily for motor reasons, the subsequent elaboration of the content component must have involved strong perceptual motivations.

**Greenberg,** although not denying the possibility of an ingestive origin of the basic syllabic pattern and adopting a sensorimotor perspective, puts considerable emphasis on perceptual factors in the evolution of syllable patterns. For example, he argues that syllables of long duration have perceptually distinctive status. Most generally, he cites evidence from vision, audition, memory, and motor function for a "general sensori-motor and information retrieval integration time constant" of 5–6 cycles per second. However, the fact that sign language production seems to have a sign output rate that is about half this value (Klima & Bellugi 1979) casts doubt of the universality of such a time constant.

**Andrew** suggests an alternative to an ingestive origins scenario buttressed by several interesting facts about mammal and bird calls. Agreeing that jaw movements are controlled by neural mechanisms that first evolved to allow biting, he also suggests that respiratory mechanisms that call for inspiratory opening and expiratory closing of the mouth may also have had a fundamental role in the origin of vocalization. From a single vocalization calling for mouth opening there may have evolved repetitive vocalizations with repetitive mouth openings, without the need for borrowing from cyclical ingestive underpinnings. Andrew believes that a similar sequence of events, though without a respiratory underpinning, might have occurred in primates in the evolution of rhythmic lipsmacking as a form of ritualization from temporally less regular oral grooming movements.

This is an interesting alternative to one that puts the main emphasis on exaptation of the cycle as such, from ingestive movements. My feeling is that the high degree of rhythmicity of the frame cycle in modern infants from the beginning of babbling onwards suggests a fundamental role of the cycle as such in speech evolution. However, the most important thing here is to have these alternatives on the table for further consideration.

**Andrew** also draws attention to the possible evolutionary importance of auditory feedback in the animal's regulation of its own call. **Abry et al.** regard this self regulatory capability as a possible first step in the evolution of the lateral extrinsic system in humans. Andrew takes issue with my conclusion, however, that vocal learning capability is negligible in other primates, citing the positive results obtained by Masataki and Fujita (1989) in a cross-fostering experiment. But as Hauser (1996) points out, "an identical study by Owren and colleagues (1992; 1993) which was ongoing at the time, has failed to provide support for the conclusions reached by Masataki and Fujita" (p. 315). Hauser concludes that "the available data on vocal production provide only weak support for the role of experience in modifying call structure" (p. 315).

**Lieberman** concludes that because dogs bark repetitively but do not have language I must be wrong about the fundamental status of the syllable in human language. Mercifully, **Andrew**'s measured analysis of the status of repetitive vocalizations in other species makes it unnecessary for me to attempt to account for the absence of language in dogs. It also seems unnecessary for me to abandon the view that the syllable is fundamental because **Lieberman, Jürgens**, and **Ohala** have pointed out that not all modern syllables involve an open-close alternation. It is interesting to note that Lieberman makes no claim for the importance of the emergence of the two-tube vocal tract in response to a target article on the evolution of speech organization. Instead, he even seems to be hinting at its irrelevance by noting that "Acoustic analysis of primate vocalizations . . . shows formant frequency transitions similar to those that convey the consonant-vowel distinctions of human speech."

**Ohala** favors a solely perceptual motivation for the initial evolution of the sound structure of speech, believing that "what are called 'syllables' are just epiphenomenal consequences of the necessity of making a succession of auditorily robust modulations in one or more acoustic parameters." One might ask why the causal burden needs to be heaped on one of the two variables – motor or perceptual – given **Lindblom's** persuasive arguments for their necessary complementarity. Ohala gives an example in which a speaker might start out with a closed vocal tract for a bilabial stop. After this "an auditorily detectable modulation" might be made by opening the tract somewhat to make a lateral such as [l]. If one takes this particular example seriously, one might wonder why a communicating hominid would ever have stop consonants as separate entities because during the closing phase there would be either some low-frequency voicing-related signal for a short time, followed by silence, or silence throughout. The voicing would not contribute place of articulation information and the silence would contribute nothing. As Ohala well knows, most of the information in stop consonants comes from making or breaking them, and so it would seem more plausible to start out with a cyclical maneuver from which the auditory correlates of making and breaking automatically fall out, rather than to construct a sound sequence by beginning from a closing phase that contributes virtually no auditory information.

Maddison and Precoda (1990) made an analysis of patterns of consonant-vowel co-occurrences in five languages in which they evaluated both perceptual and motor motivations for the co-occurrence patterns: (1) They would tend to maximize serial perceptual contrasts, which might be expected from Ohala's origins scenario. (2) They would tend towards patterns that would minimize the articulatory effort of moving from consonant to vowel targets, which would be expected from the F/C perspective. They did not find evidence for either possibility. However a reanalysis of their data combined with the data from another similar study of five different languages by Janson (1986) by Davis


https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

and I (MacNeilage & Davis 1993) reported evidence for the second trend. As noted above, CV co-occurrence patterns in babbling and infant speech also favor the second possibility. If the origin of sound sequences was primarily related to perceptual distinctiveness, one would not expect either adult speech or the earliest speechlike behavior to primarily exhibit articulatory compatibility between consonants and the vowels that followed them. And one would not expect such articulatory compatibility to be produced so fluently from the beginning of babbling if first words spoken by hominids resulted from ad hoc attempts to produce sound sequences with serial perceptual contrasts.

## R4. A medial system for frames and a lateral system for content

**R4.1. A medial system for frames.** There was more disagreement with the proposition that the frame component of speech is mediated by a medial intrinsic cortical system than with any other aspect of the target article. **Jürgens, Abbs & DePaul,** and **Lund** all strongly disagreed with this contention, concluding that the main cortical locus of both frame and content is probably lateral premotor cortex.

In my view, the fact that electrical stimulation and irritative lesions of the SMA produce rhythmic syllable repetition suggested that the SMA was involved in frame generation in humans. Part of **Jürgens**'s negative conclusions regarding a role for medial cortex in frame generation comes from his (uncontroversial) assumption that electrical stimulation is unlikely to involve normal behavior of an area. But what of the irritative lesion effects that, though also unnatural, are unnatural in a different way. It is necessary to explain why both influences on the SMA have identical effects, in the form of repetitive syllabic vocalizations, and why no such effects are obtained from lateral frontal cortex. Although these findings do not make it possible to say that the SMA is involved in frame representation, it remains possible to say that it is involved in the control of frames, in a way that is different from lateral premotor cortex.

An important clue as to how the SMA is involved in frame control comes from a recent paper by Ziegler et al. (1997) in which they report what is, to my knowledge, the first experimental study of speech in a patient with a medial cortical lesions. This patient suffered an infarct of the left anterior cerebral artery resulting in a primarily subcortical lesion which disconnected the SMA from lateral frontal cortex. Technically then, the patient could be described as a transcortical motor aphasic. This patient showed a considerable increase in speech onset latencies as a function of the number of syllables in the word to be produced from memory, most obvious when the words was a pseudoword. In contrast, intersyllable intervals during word production did not increase with number of syllables. In addition the segmental complexity of the words was not a factor in production in the way it typically would have been in a patient with a lateral premotor lesion. Ziegler et al. concluded that the role of the SMA may be in the stage of downloading of the plan for utterance syllabification.

A surprising thing about **Jürgens**'s commentary was that even though he has been most responsible for showing that the medial cortex is the main cortical control center for control of vocal communication in nonhuman primates, he gives it no role in speech, apparently concluding that evolution of speech involves only the emergence of articulatory control in lateral cortex. In the target article, I regarded such a step as saltational and was at pains to show that medial cortex did not lose a role in vocal communication with the evolution of speech. Another surprising aspect of Jürgens's commentary is his conclusion that the sharing of muscles, motoneurons, and motorcortical areas by speech and mastication is "trivial." This is tantamount to concluding that descent with modification is a trivial concept. In contrast to this biologically unorthodox stance, **Walker** provides a number of examples of the extreme conservatism of vertebrate motor evolution and emphasizes the importance of recognizing this conservatism in the construction of explanatory accounts.

**Jürgens** also purports to provide evidence that the brain stem pattern generator for mastication is unlikely to be used for speech. He reports that electrical stimulation of this pattern generator during the production of "cackling" calls in squirrel monkeys had no effects, even through such calls have a 10 Hz rhythm "with involvement of supralaryngeal muscle activity in the rhythm of the repetition rate." However inspection of spectrograms of such calls in papers by Winter et al. (1966), Schott (1975), and Jürgens (1979a) shows no obvious evidence for supralaryngeal modulation. The rhythmic component shows a harmonic structure suggesting a laryngeal source. Nevertheless Jürgens has identified a means of testing one aspect of the frame/content theory. If electrical stimulation of this region did not disturb the production of primate communicative gestures characterized by mandibular cyclicity, this would be major counterevidence to the theory.

In **Jürgens**'s opinion, the contrast between the absence of disturbances of speech rhythm following medial damage and the presence of such disturbance following lateral frontal damage is further evidence against a medial frame component. I am not aware of any quantitative data directly bearing on rhythmic disturbances in medial cortex. But Ziegler et al.'s finding that intersyllable interval did not vary with number of syllables in the word in their patient suggests that within-word rhythmic organization may be preserved in medial patients. Perhaps it depends on what one means by rhythm. It is well known (and Ziegler et al. noted this in their patient) that the speech of medial patients is marked by hesitations and word repetitions. These phenomena suggest that the downloading deficit Ziegler et al. claim to be present in medial patients has its main effect on spontaneous speech – a rhythmic effect – at the interword level. The equivalent of longer onset latencies for words with more syllables in single word production might be the interword hesitations that occur in spontaneous speech.

Contrary to **Jürgens**'s belief, Roland et al.'s (1980) paradigm involving repetitive finger movement, which yielded no SMA activity is not a manual parallel to reiterant speech. As I pointed out, in addition to multiple syllable production, reiterant speech also involves placement of stresses on the correct syllables of the words being simulated, which would, in turn, involve keeping track of the number of syllables in each word. To achieve this result, a speaker would presumably have to activate the downloading capacity suggested for the SMA by Ziegler et al. As **Abry et al.** point out, Murphy et al. (1997) have shown a dissociation between PET responses in medial but not lateral cortex associated with respiratory and phonatory aspects of speech,



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

and in lateral but not medial cortex associated with articulation. This result is obviously not consistent with Jürgens's claim of the lack of a role of medial cortex in modern speech production.

**Jürgens** and **Ohala** note differences in the role of the mandible in chewing and speech, and consider them counterexamples to the theory. Such differences are not necessarily damaging to the claim that the syllable cycle evolved from the ingestive cycle. As **Lindblom** notes, any borrowing of the ingestive cycles for speech purposes must have have been followed by shaping in terms of the needs of the new adaptation. Nevertheless, it would be useful to understand why the observed differences between the operating characteristics of two systems exist.

**Abbs & DePaul** conclude that I have fatally oversimplified my treatment of both the medial and lateral motor cortical fields. They raise a crucial issue. My task was to try to put together an enormous body of diverse information, not only on normal primate vocal communication and its ontogeny and phylogeny but also on comparative neurobiology: information of a tidy sort in the case of nonhuman invasive procedures, but of a rather untidy sort when human neuropathology was involved. Did I fail where the question of the intrinsic and extrinsic systems was involved? My response is oversimplified? Yes, necessarily: fatally oversimplified? No. Abbs & DePaul allow themselves the luxury of focussing primarily on the results of invasive studies of nonhuman primates, and in this context, interpretations based on human neuropathology are certainly crude, but necessary for any full story to emerge.

**Abbs & DePaul** take the syllabic vocalization that results from electrical stimulation in epileptic patients to be a residual of primitive vocalization in monkeys and human infants, thus supporting my thesis of a phylogenetic and ontogenetic role of this region, but denying its importance in adult speech. But the results of Ziegler et al. suggest a continuing role of the SMA in adults in the form of a downloading capacity. More generally in a summary of a recent workship on SMA function, the group concluded that the relative roles of the medial and lateral systems are very much like those that originally suggested by Goldberg (1985): "the lateral PMC may be more concerned with stimulus-triggered movements . . . while the SMA may be more involved in self-paced or internally generated movements." (Luders 1996, p. 485) Goldberg himself, in his commentary (**Goldberg & Brooks**) voiced no objection to my suggested role of the two systems in speech.

**Lund,** like **Jürgens** and **Abbs & DePaul,** sees no necessity for a role of medial cortex in speech on the grounds that cortical control of mandibular oscillation exists in the lateral cortex. As in the case of Jürgens's analysis, this conclusion begs the question of how ancestors with cortical control of vocalization that was primarily medial, evolved into modern hominds with primarily lateral control of speech. It also ignores the evidence for a continuing role of medial cortex in present day human speech. It may be worthwhile in this context to add the conclusion of Penfield and Roberts (1959) regarding the relevance of the SMA to language derived from their electrical stimulation studies: "So far as can be determined, there is no difference between the effects of electrical current when applied to the dominant Broca's Area, supplementary motor area, or parietotemporal region (Wernicke's Area) as regards the various alterations in speech" (p. 136).

**Abry et al.** seem to have unearthed an insuperable problem for the three commentators who conclude that both frame and content are controlled in lateral premotor cortex – **Abbs & DePaul, Jürgens,** and **Lund.** They refer to the work of Poeck et al. (1984), who distinguish two classes of global aphasics, the classical type with negligible comprehension or production of language and another type otherwise similar, but capable of uttering a single repeated syllable form, much as Broca's famous patient repeated "tantantan." As global aphasia is associated with relatively complete destruction of the perisylvian cortex of the left hemisphere, the subgroup who produced syllabic sequences could not be producing their simple frames under control of left lateral premotor cortex.

Although localization evidence in this study did not allow a distinction between these two groups of aphasics in terms of regions damaged versus regions spared, the existence of global aphasics with syllable repetition capacity provides additional evidence for the existence of a basic frame-related cyclical component as a distinct entity in the modern hominid brain. As **Levelt & Schiller** note, a "core point" of the F/C theory is that "there is a basic syllable cycle in the speech production system. That basic syllable cycle can run without the retrieval of stored syllable frames." This conclusion is of particular relevance to the connectionist modelling approach to speech error data discussed by **Harley.** Even if connectionist models can account for speech error data without using a frame-content distinction, as Harley claims, evidence that basic syllable repetition results from electrical stimulation and irritative lesions affecting the SMA, and is produced by some global aphasics, suggests that the concept of frame is necessary for the understanding of speech production.

In summary, three commentators deny the medial cortex a role in the frame component of the F/C theory. This claim cannot be sustained. The precise role of the medial system in speech may be one of a buffer, downloading frame-related information in the transitional phase between lexical access and movement control. More generally, the frame cycle itself seems to have neurological reality as an independent entity in modern hominids, a reality not crucially dependent on lateral premotor cortex.

A footnote: the comment of **Ghazanfar & Katz** that the neurobiology of the target article has a piecemeal quality to it is well taken. The same piecemeal quality remains present in this response. It must be conceded that I am talking about distributed systems and it is an oversimplification to talk in terms of single specific regions being solely responsible for particular functions. A better generalization, though still not ideal, is that one region is more responsible for some function than other parts of the system.

**R4.2. A lateral system for content.** There was less dissent regarding the suggested role of the lateral cortical system, than for the medial system in the F/C dichotomy, which is perhaps not surprising because this suggestion is more in line with the traditional view of neural organization for speech. On the positive side, strong support for the contention that different parts of lateral frontal cortex are involved in the control of cyclical mandibular movements associated with ingestion in monkeys and in humans is provided by **Sessle.** He even suggests a possible relation between the course of development of dental mechanoreceptors and the onset of babbling at around 6 or 7 months of age.



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

Beyond this, even though **Sessle** does not make specific suggestions as to the role of the lateral frontal cortex in the phylogeny of speech production, he is highly sympathetic to the proposal that it played an important role. **Lund** also finds this suggestion plausible. The fact that the two commentators most involved in the neurophysiology of posterior lateral frontal cortex both find the hypothesis regarding the phylogeny of this region for speech purposes plausible is encouraging.

In contrast, **Abry et al.** and **Lieberman** see little reason to implicate lateral premotor cortex in speech evolution. It is well known that small lesions in this region tend to have little effect on speech production (e.g., Mohr 1976). However, there is some consensus that somewhat larger lesions in this territory, though not necessarily involving primary cortex or prefrontal cortex typically impair speech production (Brown 1979; Galaburda 1982). In addition, the typical involvement of this area in imaging studies of speech production (see the target article, Fig. 2) and the strong connections of the area with temporoparietal cortex (Deacon 1992) are further reasons to believe it has an important role in speech production.

## R5. Evolution of vocal communication in other taxa

One useful role of the target article is to provide a basis for comparison of speech with vocal communication in other major taxa. **Scanlan & Rogers** and **Pepperberg** raise some interesting questions regarding the relation between vocal communication in birds and humans. Scanlan & Rogers suggest that, as in speech, there may have been an ingestive origin of one aspect of vocalization in parrots. They describe a prevocal ventro-caudad retraction of the larynx which is similar to a movement made in preparation for suction drinking. This appears to be an elegant piece of tinkering, where the animal takes advantage of the necessary complement of the intraoral pressure decrease that facilitates suction drinking – a sublaryngeal pressure increase facilitating a burst of vocal action.

**Scanlan & Rogers** also suggest a more general case of ingestive origins of vocal communication with their claim that the larynx itself first evolved to protect the airway during swallowing. As in humans, the possibility that vocal communication in birds developed, at least in part, from ingestion-related capabilities seems a plausible one.

**Pepperberg** presents a "me too" thesis regarding the speechlike capabilities that have been learned by her African Grey parrot. She claims that the assumption of a two-tubed vocal tract is necessary to account for sound production in parrots. She notes the presence of both glottal opening-closing alternations, and beak opening-closing alternations. She notes a kind of sound play which seems to imply the presence of word-level frames, in that birds systematically vary one sound in a word while leaving the rest of the word constant. She points out that parrots produce the acoustic transients associated with closure for consonants which are considered to play an important role in speech perception. She also notes a possible parallel in other birds to the proposed primate progression from lip movements of grooming in monkeys to communicative lipsmacks, in the form of expectant mouth openings of infant birds developing into begging calls.

Much has been made of the analogy between birds and humans in their possession of a vocal learning capacity. The capabilities of **Pepperberg**'s parrot certainly seem to raise a problem for the concept of universal grammar, which includes the assumption of a unique innate human phonological component underlying vocal communication. Parrots seem to do very well in learning to produce speechlike utterances in the absence of any such endowment.

**Pepperberg** concludes that "MacNeilages's hypothesis may have implications beyond mammals." The core of my hypothesis is that a syllabic frame evolved from ingestive cyclicities to serve speech. In the commentaries on birds, the frame concept was invoked metaphorically to describe the ventro-caudal movement noted by **Scanlan & Rogers** and the manipulation of segments in an otherwise intact word in the word play noted by Pepperberg. In both cases the term was used to denote similarities between the phenomena being described and the frame posited in the F/C theory. However, the frame of Scanlan & Rogers is dissimilar to the frame of F/C theory in that it does not serve as a receptacle for sounds. In the case of Pepperberg's frame, we do not know whether frame-related information has independent status in the production of parrot words in the way it apparently has in speech production. The sound substitutions of parrot word play, also sometimes seen in infants, could be produced without there being a frame structure for words in the sense that the term has been used in the target article.

One problem that I assume exists for the study of birds is that, unlike the case in humans who have many living relatives similar to their ancestral forms, there are no existing forms that are similar to ancestral forms of birds being studied. Consequently the phylogeny of bird vocalization is somewhat more difficult to address than that of speech, and therefore difficult to compare with speech.

## R6. The manual alternative for language evolution

**Rizzolatti** suggests a new and important means whereby elements of a manual gestural sign language might have evolved. The theory takes as a point of departure his discovery of "mirror neurons" in lateral premotor cortex – "neurons that discharge *both* when the monkey performs an action and when it observes another monkey performing it." The basic idea is that such neurons could potentiate signal production in the form of acts that would be icons for concepts that the animal had in mind. A great merit of his proposal is that it involves a natural way of linking entities in the signal to the concepts that they stand for, especially in the realm of verbs. I do not see any equally good alternative for making this linkage in the vocal-auditory modality. I doubt whether any highly elaborated manual precursor to spoken language ever evolved because of the difficulty of accounting for its ultimate displacement by spoken language. The ultimate ascendency of spoken language would seem to require the existence of a set of manual-vocal equivalences, the manual components of which ultimately die out. Although they agree that there never was a fully fledged manual language, **Woll & Sieratzki** present, in the form of "echo phonology" a number of specific manual-vocal equivalencies that occur in modern day sign languages, which therefore might have served as manual-vocal equivalences in evolution. The number of these that involve



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

labial consonant ([p], [m]) is of particular interest, considering that frames with labials ("pure frames") are considered to be the simplest frames in the F/C theory.

## R7. Broader implications

**Goldberg & Brooks** see a role for frames, as aspects of the dynamics of speech action, in the more general context of the dynamics of the internal language processes associated with the production of an utterance as a whole. This would certainly be consistent with the loss of utterance-level fluency in patients with medial lesions. However there is presumably nothing essential in mandibular oscillation as such in the mediation of this outcome, as fluent users of sign language seem to be as good as speakers at achieving overall utterance integration during language production.

**Carstairs-McCarthy** suggests that the syllable frame might be a literal precursor to a syntactic frame for the sentence. I am sympathetic to this possibility as an alternative to the de novo scenario for syntactic evolution favored by Chomsky and many other formal linguists. One important role of the syllable frame from an evolutionary perspective is that it involves the evolution of abstract status of the syllable-internal components of the frame – onset, nucleus, coda – by means of descent with modification of the concrete action of mandibular oscillation. A similar move is necessary in a neoDarwinian context, in order to get abstract syntactic entities from concrete ones in evolution. Although I am sympathetic with Carstairs-McCarthy's proposal because of its specificity and its adherence to a descent with modification paradigm, it is beyond my capabilities to evaluate it in detail.

The suggestion that human language could have developed through the interaction of procedural learning with a pre-existing system for production of socio-affective communication was presented by **Sieratzki & Woll.** This proposal, while plausible as part of an account of language evolution, has no implications for the question at issue here, which is how the organization of speech evolved.

**Buckingham** appropriately places the target article in a philosophical tradition of embodiment, with its roots in the work of David Hartley (1749) and Broca (1861), among others. And, as in the target article, I would emphasize the contrast between this perspective and the essentialist, or what Lakoff (1987) calls the objectivist perspective, with a lineage that includes Plato, Descartes, and Chomsky. This particular dog did not bark in the peer commentaries.

ACKNOWLEDGMENT

This paper was prepared with support from research grant No. HD-27733 from the Public Health Service.

# References

**[Note: The letters 'a' and 'r' before the author's initials refer to the target article and response article references, respectively.]**

Abbs, J. (1986) Invariance and variability in speech production: A distinction between linguistic intent and its neuromotor implementation. In: *Invariance and variability in speech processes,* ed. J. Perkell & D. Klatt. Erlbaum. [JHA]

Abbs, J. & Gracco, V. (1984) Control of complex motor gestures: Orofacial muscle responses to load perturbation of the lip during speech. *Journal of Neurophysiology* 51:705–23. [JHA]

Abbs, J. & Rosenbek, J. (1985) Some motor control perspectives on apraxia of speech and dysarthria. In: *Recent advances in speech disorders,* ed. J. Costello. College Hill Press. [JHA]

Abbs, J. & Welt, C. (1985) Structure and function of the lateral precentral cortex: Significance for speech motor control. In: *Recent advances in speech science,* ed. R. Daniloff. College Hill Press. [JHA]

Abry, C. (1997) *From speech to speech. . . . Or, from the vocal Self to the narrative Other in a "theory of mind."* Doctoral d'Etat des Lettres et Sciences Humaines, Grenoble-Stendhal. [CA]

Aich, H., Moos-Heilen, R. & Zimmerman, E. (1990) Vocalization of adult gelada baboons (*Theropithecus gelada*): Acoustic structure and behavioral context. *Folia Primatologica* 55:109–32. [aPFM]

Alexander, G. E. & Crutcher, M. D. (1990) Preparation for movement: Neural representations of intended direction in three motor areas of the monkey. *Journal of Neurophysiology* 64:133–50. [JHA]

Alexander, M. P., Naeser, M. A. & Palumbo, C. L. (1987) Correlations of subcortical CT lesion sites and aphasia profiles. *Brain* 110:961–91. [PL]

Anderson, S. (1981) Why phonology isn't natural. *Linguistic Inquiry* 12:493–539. [aPFM]

Andrew, R. J. (1963) The origins and evolution of the calls and facial expressions of the primates. *Behavior* 20:1–109. [RJA]

(1964) Vocalization in chicks and the concept of "stimulus contrast." *Animal Behaviour* 12:64–76. [RJA]

(1976) Use of formants in the grunts of baboons and other nonhuman primates. *Annals of the New York Academy of Sciences* 280:673–93. [aPFM]

Arai, T. & Greenberg, S. (1997) The temporal properties of spoken Japanese are similar to those of English. In: *Proceedings of Eurospeech, Rhodes, Greece, vol. 2,* pp.1011–14. [SG]

Arbib, M. A. & Rizzolatti, G. (1997) Neural expectations: A possible evolutionary path from manual skill to language. *Communication and Cognition* 29:393–424. [GR]

Armstrong, D., Stokoe, W. C. & Wilcox, S. (1994) *Gesture and the nature of language.* Cambridge University Press. [BW]

Baddeley, A. D. (1986) *Working memory.* Clarendon Press. [aPFM]

(1995) Working memory. In: *Principles of neuroscience,* ed. M. S. Gazzaniga. MIT Press. [aPFM]

Barlowe, S. M. & Farley, W. H. (1989) Neurophysiology of speech. In: *Neural bases of speech, hearing and language,* ed. D. P. Kuehn, M. L. Lemme & J. M. Baumgartner. College Hill Press. [LM]

Beaumont, S. L. & Bloom, K. (1993) Adults' attributions of intentionality to vocalizing infants. *First Language* 13:325–47. [KB]

Beck, P. D., Pospichal, M. W. & Kaas, J. H. (1996) Topography, architecture, and connections of somatosensory cortex in opossums - evidence for 5 somatosensory areas. *Journal of Comparative Neurology* 366:109–33. [SFW]

Beevor, C. E. & Horsley, V. (1894) A further minute analysis by electrical stimulation of the so-called motor region (facial area) of the cortex cerebri in the monkey (*Macacus sinicus*). *Philosophical Transactions of the Royal Society of London B* 185:30–81. [JPL]

Bemer, F. (1923) Physiologie nerveuse de la mastication chez le char et le lapin. Réflexes de mastication. Réponses masticatrices corticales et centre cortical du gôut. *Archives Internationales de Physiologie et Biochimie* 21:308–52. [JPL]

Blinkov, S. M. & Glezer, I. I. (1968) *The human brain in figures and tables.* Plenum Press. [AAG]

Bloom, K. (1988) Quality of adult vocalizations affects the quality of infant vocalisations. *Journal of Child Language* 15:469–80. [KB]

(1989) Duration of early vocal sounds. *Infant Behavior and Development* 12:245–50. [rPFM]

(1993) Infant voice and mouth movements affect adult attributions. *Journal of Language and Social Psychology* 12:269–87. [KB]

Bloom, K., D'Odorico, L. & Beaumont, S. (1993) Adult preferences for syllable vocalizations: Generalizations to parity and native language. *Infant Behavior and Development* 16:109–20. [KB]

Bloom, K. & Lo, E. (1990) Adult perceptions of vocalizing infants. *Infant Behavior and Development* 13:209–19. [KB]

Bloom, K. & Masataki, N. (1996) Japanese and Canadian impressions of vocalizing infants. *International Journal of Behavioral Development* 19:89–99. [KB]

Bloom, K., Russell, A. & Wassenberg, K. (1987) Turn taking affects the quality of infant vocalizations. *Journal of Child Language* 14:211–27. [KB]

Blumstein, S. E., Baker, H. & Goodglass, H. (1977) Phonological factors in auditory comprehension in aphasia. *Neuropsychologia* 15:19–30. [aPFM]

Bouabana, S. & Maeda, S. (1998) Multipulse LPC modeling of articulatory movements. *Speech Communication* (in press). [SG]

Bouillaud, J. B (1825) Recherches cliniques propres a demonstrer que la perte de la parole correspond à la lesion des lobles anterieurs du cerveau. Et a confirmer l'opinion de M. Gall aur le seige de l'orange du language articule. *Archives Generales de Medecine* 8:25–45. [HWB]



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

Brickner, R. M. (1940) A human cortical area producing repetitive phenomena when stimulated. *Journal of Neurophysiology* 3:128–30. [aPFM]
Brittan-Powell, E. F., Dooling, R. J. & Farabaugh, S. M. (1997) Vocal development in budgerigars (*Melopsittacus undulatus)* contact calls. *Journal of Comparative Psychology* 111:226–41. [IMP]
Broad, D. J. (1973) Phonation. In: *Normal aspects of speech, hearing and language,* ed. F. D. Minifie, T. J. Hixon & F. Williams. Prentice-Hall. [aPFM]
Broca, P (1861/1960) Remarks on the seat of the faculty of articulate language, followed by an observation of aphemia. In: *Some papers on the cerebral cortex,* ed. G. von Bonin. C. C. Thomas. [HWB, rPFM]
Browman, C. P. & Goldstein, L. (1986) Toward an articulatory phonology. In: *Phonology Yearbook 3.* Cambridge University Press. [arPFM]
(1992) Articulatory phonology: An overview. *Phonetica* 49:155–80. [LMe]
(1995) Gestural syllable position effects in American English. In: *Producing speech: Contemporary issues,* ed. F. Bell-Berti & L. J. Raphael. AIP Press. [LMe]
Brown, J. W. (1979) Language representation in the brain. In: *Neurobiology of social communication in primates,* ed. H. D. Steklis & M. J. Raleigh. Academic Press. [rPFM]
Buckingham, H. & Yule, G. (1987) Phonemic false evaluation: Theoretical and clinical aspects. *Clinical Linguistics and Phonetics* 1:113–25. [KGM]
Carey, D. P., Perrett, D. J. & Oram, M. W. (1997) Recognizing, understanding and reproducing action. In: *Handbook of neuropsychology, vol. 11,* ed. F. Boller & J. Grafman. Elsevier. [GR]
Carré, R., Lindblom, B. & MacNeilage, P. F. (1995) Acoustic factors in the evolution of the vocal tract. (Translation) *C. R. Academie des Sciences Paris* t 320, Series IIb, 471–76. [aPFM]
Carstairs-McCarthy, A. (1998) Synonymy avoidance, phonology and the origin of syntax. In: *Approaches to the evolution of language: Social and cognitive bases,* ed. J. R. Hurford, M. Studdert-Kennedy & C. Knight. Cambridge University Press. [AC- M]
(forthcoming) *The origins of complex language: An inquiry into the evolutionary beginnings of sentences, syllables and truth.* Oxford University Press. [AC-M]
Chauvel, P. C. (1976) *Les stimulations de l'aire motrice supplementaire chez l'homme.* Theses, Unversité de Rennes. [aPFM]
Chomsky, N. (1986) *Knowledge of language: Its nature, origin, and use.* Praeger. [aPFM]
(1988) *Language and problems of knowledge: The Managua lectures.* MIT Press. [aPFM]
Chomsky N. & Halle, M. (1968) *The sound pattern of English.* Harper and Row. [aPFM]
Cohen, A. H. (1988) Evolution of the vertebrate central pattern generator for locomotion. In: *Neural control of rhythmic movements,* ed. A. Cohen, S. Rossignol & S. Grillner. Wiley. [aPFM]
Corballis, M. C. (1992) On the evolution of language and generativity. *Cognition* 44:197–226. [GR]
Damasio, A. (1985) Understanding the mind's will. *Behavioral and Brain Sciences* 8:589. [JHA]
Damasio, A. & Damasio, H. (1992) Brain and language. *Scientific American* 267:63–71. [aPFM]
Damasio, A. R. & Geschwind, N. (1984) The neural basis of language. *Annual Review of Neuroscience* 7:127–47. [aPFM]
Darley, F. L., Aronson, A. E. & Brown, J. R. (1975) *Motor speech disorders.* Saunders. [UJ]
Darwin, C. (1959) *The origin of species.* John Murray. [aPFM]
(1871) *The descent of man.* Great Books, Encyclopedia Britannica. [aPFM]
Davis, B. L. & MacNeilage, P. F. (1995) The articulatory basis of babbling. *Journal of Speech and Hearing Research* 38:1199–211. [LM, arPFM, JP]
Davis, P. J., Zhang, S. P., Winkworth, A. & Bandler, R. (1996) Neural control of vocalization: Respiratory and emotional influences. *Journal of Voice* 10:23–38. [SFW]
Day, L. B. & MacNeilage, P. F. (1996) Postural asymmetries and language lateralization in humans (*Homo sapiens). Journal of Comparative Psychology* 110:86–96. [aPFM]
Deacon, T. W. (1992) The neural circuits underlying primate calls and human language. In: *Language origin: A multidisciplinary approach,* ed. J. Wind, B. Chiarelli, B. Bichakjian & A. Nocentini. Kluwer. [arPFM]
Decety, J., Grèzes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E., Grassi, F. & Fazio, F. (1997) Brain activity during observation of actions. Influence of action content and subject's strategy. *Brain* 120:1763–77. [GR]
Deiber, M., Passingham, R., Colebatch, J., Friston, K., Nixon, P. & Frackowiak, R. (1991) Cortical areas and the selection of movement: A study with positron emission tomography. *Experimental Brain Research* 84:393–402. [JHA]
De Jong, K., Beckman, M. E. & Edwards, J. (1993) The interplay between prosodic structure and coarticulation. *Language and Speech* 36:197–212. [LMe]
Dell, G. S. (1986) A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93:283–321. [GG, TAH, aPFM]
(1988) The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language* 27:124–42. [WJML]
Dell, G. S., Juliano, C. & Govindjee, A. (1993) Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science* 17:149–95. [TAH]
Demonet, J-F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J-L., Wise, R., Rascol, A. & Frackowiak, R. S. J. (1992) The anatomy of phonological and semantic processing in normal subjects. *Brain* 115:1753–68. [aPFM]
Demonet, J. F., Wise, R. & Frackowiak, R. S. J. (1993) Language functions explored in normal subjects by positron emission tomography: A critical review. *Human Brain Mapping* 1:39–47. [aPFM]
Dettmers, C., Fink, G., Lemon, R., Stephan, K., Passingham, R. Silbersweig, D., Holmes, A., Ridding, M., Brooks, D. & Frackowiak, R. (1995) Relation between cerebral activity and force in the motor areas of the human brain. *Journal of Neurophysiology* 74(2):802–15. [JHA]
De Voogd, T. J., Krebs, J. R., Healy, S. D. & Purvis, A. (1993) Relations between song repertoire size and the volume of brain nuclei related to song: Comparative evolutionary analysis among oscine birds. *Proceedings of the Royal Society, London, B* 254:75–82. [IMP]
Di Benedetto, M.-G. (1989) Frequency and time variations of the first formant: Properties relevant to the perception of vowel height. *Journal of the Acoustical Society of America* 86:67–77. [JP]
(1994) Acoustic and perceptual evidence of a complex relation between $F_1$ and $F_0$ in determining vowel height. *Journal of Phonetics* 22:205–24. [JP]
Dick, J. P. R., Cantello, R., Buruma, O., Gioux, M., Benecke, R., Day, B. L. et al. (1987) The Bereitschaftspotential, L-DOPA, and Parkinson's disease. *Electroencephalography and Clinical Neurophysiology* 66:263–74. [GG]
Dinner, D. S. & Luders, H. O. (1995) Human supplementary sensorimotor area: Electrical stimulation and movement-related potential studies. In: *Epilepsy and the functional anatomy of the frontal lobe,* ed. H. H. Jasper, S. Riggio & P. S. Goldman-Rakic. Raven Press. [aPFM]
Dinner, D. S., Luders, H., Morris, H., Wylie, E. & Kramer, R. (1987) Human supplementary motor area (SMA): Electrical stimulation study. *Epilepsia* 28:619. [JHA]
Donald, M. (1991) *Origins of the modern mind: Three stages in the evolution of culture and cognition.* Harvard University Press. [aPFM]
Dooling, R. J., Best, C. T. & Brown, S. D. (1995) Discrimination of synthetic full-formant and sinewave /ra-la/ continua by budgerigars (*Melopsittacus undulatus)* and zebra finches (*Taeniopygia guttata). Journal of the Acoustical Society of America* 97:1839–46. [IMP]
Dooling, R. J., Okanoya, K. & Brown, S. D. (1989) Speech perception by budgerigars (*Melopsittacus undulatus)*: The voiced-voiceless distinction. *Perception and Psychophysics* 46:65–71. [IMP]
Dressnandt, J. & Jürgens, U. (1992) Brain stimulation-induced changes of phonation in the squirrel monkey. *Experimental Brain Research* 89:549–59. [UJ]
Dunbar, R. I. M. & Dunbar, P. (1975) Social dynamics of gelada baboons. In: *Contributions to primatology, vol. 6.* Karger. [aPFM]
Durand, S. E., Heaton, J. T., Amateau, S. K. & Brauth, S. E. (1997) Vocal control pathways through the anterior forebrain of a parrot (*Melopsittacus undulatus). Journal of Comparative Neurology* 377:179–206. [IMP]
Ebrahimi, A., Pochet, R. & Roger, M. (1992) Topographical organization of the projections from physiologically identified areas of the motor cortex to the striatum in the rat. *Neuroscience Research* 14:39–60. [SFW]
Eccles, J. C. (1982) The initiation of voluntary movements by the supplementary motor area. *Archiv für Psychiatrie und Nervenkrankheiten* 231:423–41. [UJ, aPFM]
Eddins, D. A. & Green, D. M. (1995) Temporal integration and temporal resolution. In: *Hearing. Handbook of perception and cognition, 2nd edition,* ed. B. C. J. Moore. Academic Press. [SG]
Elias, L. J. & Bryden, M. P. (1997) Footedness is a better predictor of language lateralization than handedness. *Laterality* 3:41–52 [aPFM]
Elman, J. L. (1990) Finding structure in time. *Cognitive Science* 14:213–52. [TAH]
Enomoto, S., Kohase, H. & Nakamura, Y. (1995) Dual brain-stem projection from the cortical masticatory area in guinea-pig. *Neuroreport* 6:1573–77. [SFW]
Erickson, D., Lenzo, K. & Fujimura, O. (1994) Manifestations of contrastive emphasis in jaw movement. *Journal of the Acoustical Society of America* 95:2822. [rPFM]
Erickson, T. C. & Woolsey, C. N. (1951) Observations on the supplementary motor area of man. *Transactions of the American Neurological Association* 76:50–52. [aPFM]
Evarts, E. V. (1982) Analogies between central motor programs for speech and for limb movements. In: *Speech motor control,* ed. S. Grillner, B. Lindblom, J. Lubker & A. Persson. Pergamon Press. [LM]
Fadiga, L., Fogassi, L., Pavesi, G. & Rizzolatti, G. (1995) Motor facilitation during



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

action observation: A magnetic stimulation study. *Journal of Neurophysiology* 73:2608–11. [GR]
Falk, D. (1983) Cerebral cortices of East African early hominids. *Science* 222:1072–74. [GR]
Fant, C. G. M. (1960) *Acoustic theory of speech production.* Mouton. [aPFM]
Fiez, J. A., Raichle, M. E., Tallal, P. A. & Petersen, S. E. (1993) Activation of a left frontal area near Broca's areas during auditory detection and phonological access tasks. *Journal of Cerebral Blood Flow and Metabolism 1*3 (suppl. 1):S519. [aPFM]
Fink, G., Frackowiak, R., Pietrzyk, U. & Passingham, R. (1996) Multiple nonprimary motor areas in the human cortex. *Journal of Neurophysiology* 77:2164–74. [JHA]
Foerster, O. (1931) The cerebral cortex in man. *Lancet* 221:309–12. [JHA]
(1936a) The motor cortex in man in the light of Hughlings Jackson's doctrines. *Brain* 59:135–59. [aPFM]
(1936b) Motorische Felder und Bahnen. In: *Handbuch der Neurobiologie, Band IV,* ed. O. Bumke & O. Foerster. Springer. [UJ]
Fox, P. T. (1995) Broca's area: Motor encoding in somatic space. *Behavioral and Brain Sciences* 18:344–45. [aPFM]
Fox, P. T., Ingram, R., Ingram, J., Hirsch, T., Hunter-Downs, J., Martin, C., Jerabek, P., Glass, T. & Lancaster, J. (1996) A PET study of the neural system of stuttering. *Nature* 382:158–62. [JHA]
Fox, P. T., Mikiten, S., Davis, G. & Lancaster, J. L. (1995) Brainmap: A database of human functional brain mapping. In: *Advances in functional neuroimaging: Technical foundations,* ed. R. W. Thatcher, M. Hallet & T. Zeffiro. Academic Press. [aPFM]
Freedman, M., Alexander, M. P. & Naeser, M. A. (1984) Anatomic basis of transcortical motor aphasia. *Neurology* 34:409–17. [aPFM]
Fried, I. (1996) Electrical stimulation of the supplementary sensorimotor area. In: *Advances in neurology, vol. 70: Supplementary motor area,* ed. H. Luders. Lippincott-Raven. [JHA]
Fried, I., Katz, A., McCarthy, G., Sass, K., Williamson, P., Spenser, S. & Spenser, D. (1991) Functional organization of human supplementary motor cortex studied by electrical stimulation. *Journal of Neuroscience* 11:3656–66. [JHA]
Frith, C. D. (1992) *The cognitive neuropsychology of schizophrenia.* Erlbaum. [CA]
Frith, C. D., Friston, K. J., Liddle, P. F. & Frackowiak, R. S. J. (1991) A PET study of word finding. *Neuropsychologia* 29:1137–48. [aPFM]
Froeschels, E. (1951) Chewing method as therapy. *Archives of Otolaryngology* 56:427–34. [JJO]
Fromkin, V. A. (1971) The non-anomalous nature of anomalous utterances. *Language* 47:27–52. [WJML]
Fromkin, V. A., ed. (1973) *Speech errors as linguistic evidence.* Mouton. [aPFM]
Galaburda, A. M. (1982) Histology, architectonics and asymmetry of language areas. In: *Neural models of language processes,* ed. M. A. Arbib, D. Caplan & J. Marshall. Academic Press. [rPFM]
Galaburda, A. M. & Pandya, D. (1982) Role of architectonics and connections in the study of primate brain evolution. In: *Primate brain evolution,* ed. E. Armstrong & D. Falk. Plenum Press. [JHA]
Gallese, V., Fadiga, L., Fogassi, L. & Rizzolatti, G. (1996) Action recognition in the premotor cortex. *Brain 1*19:593–609. [GR]
Garrett, M. F. (1975) The analysis of sentence production. In: *The psychology of learning and motivation, vol. 9,* ed. G. H. Bower. Academic Press. [TAH, WJML]
(1988) Processes in language production. In: *Linguistics: The Cambridge survey, vol. III: Language: Psychological and biological aspects,* ed. F. J. Newmeyer. Cambridge University Press. [AC-M, aPFM]
Gentilucci, M., Fogassi, L., Luppino, G., Matelli, M., Camarda, R. & Rizzolatti, G. (1988) Functional organization of inferior area 6 in the macaque monkey. I. Somatotopy and the control of proximal movements. *Experimental Brain Research* 71:475–90. [aPFM]
Geschwind, N., Quadfassel, F. A. & Segarra, J. (1968) Isolation of the speech area. *Neuropsychologia* 6:327–40. [aPFM]
Glenberg, A. M. (1997) What memory is for. *Behavioral and Brain Sciences* 20:1–55. [HWB]
Godfrey, J. J., Holliman, E. C. & McDaniel, J. (1992) SWTCHBOARD: Telephone speech corpus for research and development. In: *ICASSP-92 IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 1,* pp. 517–20. [SG]
Goldberg, G. (1985) Supplementary motor area structure and function: Review and hypothesis. *Behavioral and Brain Sciences* 8:567–616. [arPFM]
(1992) Premotor systems, attention to action and behavioral choice. In: *Neurobiology of motor program selection,* ed. J. Kein, C. R. McCrohan & W. Winlow. Pergamon Press. [aPFM]
Gottschalk, M., Lemon, R. & Kuypers, H. (1983) Afferent and efferent connections of the postarcuate region of the monkey cerebral cortex. *Society for Neuroscience Abstracts* 9:490. [JHA]
Gould, S. J. (1977) *Ontogeny and phylogeny.* Bellknap. [aPFM]
Gracco, V. & Abbs, J. (1989) Sensorimotor characteristics of speech motor sequences. *Experimental Brain Research* 75:586–98. [JHA]
Grafton, S. T., Arbib, M. A., Fadiga, L. & Rizzolatti, G. (1996) Localization of grasp representation in humans by PET: 2. Observation compared with imagination. *Experimental Brain Research* 112:103–11. [GR]
Granit, R. (1977) *The purposive brain.* MIT Press. [BL]
Grau, S. M., Robb, M. P. & Cacace, A. T. (1995) Acoustic correlates of inspiratory phonation during infant cry. *Journal of Speech and Hearing Research* 38:373–81. [JJO]
Green, S. (1975) Variations of vocal pattern with social situation in the Japanese monkey (*Macaca fuscata*): A field study. In: *Primate behavior, vol. 4: Developments in field and laboratory research,* ed. L. A. Rosenblum. Academy Press. [aPFM]
Greenberg, S. (1996) Understanding speech understanding - towards a unified theory of speech perception. In: *Proceedings of the ESCA Tutorial and Advanced Research Workshop on the Auditory Basis of Speech Perception, Keele, England,* pp. 1–8. [SG]
(1997) On the origins of speech intelligibility in the real world. In: *Proceedings of the ESCA Workshop on Robust Speech Recognition for Unknown Communication Channels, Pont-à-Mousson, France,* pp. 23–32. [SG]
Greenberg, S., Hollenback, J. & Ellis, D. (1996) Insights into spoken language gleaned from phonetic transcription of the Switchboard corpus. In: *Proceedings of the Fourth International Conference on Spoken Language (ICSLP), Philadelphia,* pp. S24–27. [SG]
Greenberg, S. & Kingsbury, B. (1997) The modulation spectrogram: In pursuit of an invariant representation of speech. In: *ICASSP-97 IEEE International Conference on Acoustics, Speech and Signal Processing, Munich, vol. 3,* pp. 1647–50. [SG]
Grillner, S. (1985) Neurological basis of rhythmic motor acts in vertebrates. *Science* 228:143–49. [JPL]
Grudin, J. (1981) The organization of serial order in typing. Unpublished doctoral dissertation, University of California at San Diego. [aPFM]
Haeckel, E. (1896) *The evolution of man, Vol. 1.* D. Appleton. [aPFM]
Hall, W. S., Brauth, S. E. & Heaton, J. T. (1994) Comparison of the effects of lesions in nucleus basalis and Field L on vocal learning in the budgerigar. *Brain Behavior and Evolution* 44:133–48. [RJA]
Hamdy, S., Aziz, Q., Rothwell, J. C., Hobson, A., Barlow, J. & Thompson, D. G. (1997) Cranial nerve modulation of human cortical swallowing motor pathways. *American Journal of Physiology - Gastrointestinal and Liver Physiology* 35:G802-G808. [SFW]
Harley, T. A. (1993) Phonological activation of semantic competitors during lexical access in speech production. *Language and Cognitive Processes* 8:291–309. [TAH]
Harley, T. A. & MacAndrew, S. B. G. (1995) Interactive models of lexicalisation: Some constraints from speech error, picture naming, and neuropsychological data. In: *Connectionist models of memory and language,* ed. J. P. Levy, D. Bairaktaris, J. A. Bullinaria & P. Cairns. UCL Press. [TAH]
Hartley, D. (1749) *Observations on man, his frame, his duty, and his expectations.* Reprinted 1976, Scholar's Fascimiles and Reprints. [HWB, rPFM]
Hauser, M. D. (1996) *The evolution of communication.* MIT Press. [rPFM]
Hauser, M. D., Evans, C. S. & Marler, P. (1993) The role of articulation in the production of rhesus monkey, *Macaca mulatta,* vocalizations. *Animal Behaviour* 45:423–33. [AAG, JS]
Hawkins, S. (1984) On the development of motor control in speech: Evidence from studies of temporal coordination. In: *Speech and language: Advances in basic research and practice, 11,* ed. N. J. Lass. Academic Press. [LMe]
Hayes, K. J. & Nissen, C. H. (1971) Higher mental functions of a home-raised chimpanzee. In: *Behaviour of nonhuman primates, vol. 4,* ed. A. M. Schrier & F. Stollnitz. Academic Press. [SFW]
He, S., Dum, R. & Strick, P. (1995) Topographic organization of corticospinal projections from the frontal lobe: Motor areas on the medial surface of the hemisphere. *Journal of Neuroscience* 15:3284–306. [JHA]
Hellige, J. B. (1993) *Hemispheric asymmetry: What's right and what's left.* Harvard University Press. [JSS]
Hixon, T. J. (1973) Respiratory function in speech. In: *Normal aspects of speech, hearing and language,* ed. F. D. Minifie, T. J. Hixon & F. Williams. Prentice-Hall. [aPFM]
Hofsten, C. von. (1984) Developmental changes in the organization of prereaching movements. *Developmental Psychology* 20:378–88. [aPFM]
(1986) The early development of the manual system. In: *Precursors to infant speech,* ed. B. Lindblom & R. Zetterstrom. Stockton Press. [aPFM]
Holloway, R. L. (1985) The past, present, and future significance of the lunate sulcus in early hominid evolution. In: *Hominid evolution: Past, present, and future,* ed. P. V. Tobias. Liss. [GR]
Homberger, D. G. (1980) Funktionell-morphologische Untersuchungen zur


https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

Radiation der ernahrungs-und Trinkmethoden der Papageien. *Bonner Zoologische Monographien* 13 [JS]
(1986) *The lingual apparatus of the African grey parrot,* Psittacus erithacus Linne *(Aves: Psittacidae) description and theoretical mechanical analysis. Ornithological Monographs, No. 39.* The American Ornithologists' Union. [IMP]
Huang, C.-S., Hiraba, H., Murray, G. M. & Sessle, B. J. (1989) Topographical distribution and functional properties of cortically-induced rhythmical jaw movements in the monkey (*Macaca fascicularis). Journal of Neurophysiology* 61:635–50. [AAG, JPL, BJS]
Huard, H., Lund, J. P. & Dubuc, R. (1995) A study of trigeminal premotor neurons in lampreys. *Society for Neuroscience Abstracts* 21:142. [JPL]
Humphrey, N. (1986) *The inner eye.* Faber & Faber. [BL]
Hura, S. L., Lindblom, B. & Diehl, R. L. (1994) On the role of perception in shaping phonological assimilation rules. *Language and Speech* 35:59–72. [aPFM]
Huxley, T. H. (1917) *Methods and results: Essays.* Appleton and Co. [aPFM]
Iriki, A., Nozaki, S. & Nakamura, Y. (1988) Feeding behavior in mammals: Corticobulbar projection is recognized during conversion from sucking to chewing. *Developmental Brain Research* 44:189–96. [BJS]
Itoh, M. & Sasanuma, S. (1984) Articulatory movements in apraxia of speech. In: *Apraxia of speech: Physiology, acoustics, linguistics, and management,* ed. J. Rosenbek, M. McNeil & A. Aronson. College Hill Press. [KGM]
Jackendoff, R. (1977) *X syntax: A study of phrase structure.* MIT Press. [AC-M]
Jacob, F. (1977) Evolution and tinkering. *Science* 196:1161–66. [aPFM]
Jahanshahi, M., Jenkins, I. H., Brown, R. G., Marsden, C. D., Passingham, R. E. & Brooks, D. J. (1995) Self-initiated versus externally-triggered movements. I. An investigation using measurement of regional cerebral blood flow with PET and movement-related potentials in normal and Parkinson's disease subjects. *Brain* 118:913–33. [GG]
Jakobson, R. (1940) Kindersprache, Aphasie und allgemeine Lautgesetze. In: *Selected writings.* Mouton.
(1968) *Child language, aphasia, and phonological universals,* trans. A. R. Keiler. Mouton. [PL, aPFM]
Jakobson, R., Fant, C. G. M. & Halle, M. (1951) *Preliminaries to speech analysis.* MIT Press. [aPFM]
Janson, T. (1986) Cross-linguistic trends in CV sequences. *Phonology Yearbook* 3:179–96. [rPFM]
John, E. R. (1967) *Mechanisms of learning and memory.* Academic Press. [SG]
Jonas, S. (1981) The supplementary motor region and speech emission. *Journal of Communication Disorders* 14:349–73. [aPFM]
Jones, J. A., Purdy, K. & Munhall, K. G. (1997) Constraints on syllable structure in speech. Paper presented at the meetings of the Acoustical Society of America, San Diego. [KGM]
Jones, S. S. (1996) Imitation or exploration? Young infants matching of adults' oral gestures. *Child Development* 67:1952–69. [KB]
Jürgens, U. (1976) Reinforcing concomitants of electrically elicited vocalizations. *Experimental Brain Research* 26:203–14. [JHA]
(1979a) Neural control of vocalization in non-human primates. In: *Neurobiology of social communication in primates,* ed. H. D. Steklis & M. J. Raleigh. Academic Press. [JHA, arPFM]
(1979b) Vocalization as an emotional indicator. A neuroethological study in the squirrel monkey. *Behaviour* 69:88–117. [JHA, UJ]
(1987) Primate communication: Signalling, vocalization. In: *Encyclopedia of Neuroscience,* ed. G. Adelman. Birkhauser. [aPFM]
(1992) On the neurobiology of vocal communication. In: *Nonverbal vocal communication: Comparative and developmental approaches,* ed. M. Papousek, H. Papousek & U. Jürgens. Cambridge University Press. [CA]
(1994) The role of the periaqueductal grey in vocal behaviour. *Behavioural Brain Research* 62:107–17. [SFW]
(1995) Neuronal control of vocal production in human and non-human primates. In: *Current topics in primate vocal communication,* ed. E. Zimmerman, J. D. Newman & U. Jurgens. Plenum Press. [aPFM]
Jürgens, U., Kirzinger, A. & von Cramon, D. (1982) The effect of deep-reaching lesions in the cortical face area on phonation: A combined case report and experimental monkey study. *Cortex* 18:125–40. [aPFM]
Jusczyk, P. W., Friederici, A. D., Wessels, J. M., Svenkerud, V. Y. & Jusczyk, A. M. (1993) Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language* 32:402–20. [LMe]
Kaye, J., Lowenstamm, J. & Vergnaud, J.-R. (1990) Constituent structure and government in phonology. *Phonology* 7:193–231. [Ac-M]
Keizer, K. & Kuypers, H. (1989) Distribution of corticospinal neurons with collaterals to the lower brain stem nuclei reticular formation in monkey (*Macaca fascicularis). Experimental Brain Research* 74:311–18. [JHA]
Kelso, J. A. S., Tuller, B., V-Bateson, E. & Fowler, C. A. (1984) Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance* 10:812–32. [AAG]
Kent, R. D. (1981) Articulatory-acoustic perspectives on speech development. In: *Language behavior in infancy and early childhood,* ed. R. E. Stark. Elsevier-North Holland. [KB]
(1993) Infants and speech: Seeking patterns. *Journal of Phonetics* 21:117–23. [LMe]
Kent, R. & Rosenbek, J. (1983) Acoustic patterns of apraxia of speech. *Journal of Speech and Hearing Research* 26:231–48. [KGM]
Kim, J. H. S., Relkin, N. R., Lee, K.-M. & Hirsch, J. (1997) Distinct cortical areas associated with native and second languages. *Nature* 388:171–74. [CA]
Kingston, J. & Macmillan, N. A. (1995) Integrality of nasalization and F1 in vowels in isolation and before oral and nasal consonants: A detection-theoretic application of the Garner paradigm. *Journal of the Acoustical Society of America* 97:1261–85. [JP]
Kirzinger, A. & Jürgens, U. (1982) Cortical lesion effects and vocalization in the squirrel monkey. *Brain Research* 233:299–315. [JHA]
(1994) Role of extralaryngeal muscles in phonation of subhuman primates. *Journal of Comparative Physiology A* 175:215–22. [UJ]
Klima, E. & Bellugi, U. (1979) *The signs of language.* Harvard University Press. [rPFM]
Kolk, H. H. J. (1995) A time-based approach to agrammatic production. *Brain and Language* 50:282–303. [GG]
Kolk, H. H. J. & van Grunsven, M. F. (1985) Agrammatism as a variable phenomenon. *Cognitive Neuropsychology* 2:347–84. [GG]
Kozhevnikov, V. A. & Chistovich, L., eds. (1965) *Speech: Articulation and perception.* Clearing House for Federal, Scientific and Technical Information, JPRS, 30:543. [aPFM]
Krubitzer, L., Manger, P., Pettigrew, J. & Calford, M. (1995) Organization of somatosensory cortex in monotremes - in search of the prototypical plan. *Journal of Cognitive Neurology* 351:261–306. [SFW]
Kuhl, P. K. & Meltzoff, A. N. (1982) The bimodal perception of speech in infancy. *Science* 218:1138–41. [aPFM]
Kurata, K. & Wise, S. P. (1988) Premotor and supplementary motor cortex in rhesus monkeys: Neuronal activity during externally- and internally-instructed motor tasks. *Experimental Brain Research* 72:237–48. [AAG]
Kypers, H. G. J. M. (1958) Some projections from the pericentral cortex to the pons and lower brain stem in the monkey and chimpanzee. *Journal of Comparative Neurology* 100:221–55. [JPL]
Lakoff, G. (1987) *Women, fire and dangerous things: What categories reveal about the mind.* The University of Chicago Press. [arPFM]
(1988) Smolensky, semantics, and the sensorimotor system. *Behavioral and Brain Sciences* 11:39–40. [HWB]
Lancaster, J. (1973) *Primate behavior and the emergence of human culture.* Holt, Rinehart Winston. [aPFM]
Lang, W., Obrig, H., Lindinger, G., Cheyne, D. & Deecke, L. (1990) Supplementary motor area activation while tapping bimanually different rhythms in musicians. *Experimental Brain Research* 79:504–14. [GG]
Lashley, K. S. (1951) The problem of serial order in behavior. In: *Cerebral mechanisms in behavior: The Hixon symposium,* ed. L. A. Jeffress. Wiley. [aPFM]
Lawson, L. (1983) Multi-channel signs. In: *Language in sign,* ed. J. G. Kyle & B. Woll. Croom Helm. [BW]
Lazar, G., Szabo, T., Libouban, S., Ravailleveron, M., Toth, B. & Brandle, K. (1992) Central projections and motor nuclei of the facial, glossopharyngeal, and vagus nerves in the mormyrid fish gnathonemus petersii. *Journal of Comparative Neurology* 325:343–58. [SFW]
LeBlanc, P. (1992) Language localization with activation PET scanning. *Journal of Neurosurgery* 31:369–73. [aPFM]
LeMay, M. & Geschwind, N. (1976) Hemispheric differences in the brains of great apes. *Brain, Behavior and Evolution* 11:48–52. [JHA]
Lenneberg, E. (1967) *Biological foundations of language.* Wiley. [JHA]
Levelt, W. J. M. (1989) *Speaking: From intention to articulation.* MIT Press. [WJML, aPFM]
(1992) Accessing words in speech production: Stages, processes and representations. *Cognition* 42:1–22. [WJML, aPFM]
Levelt, W. J. M., Roelofs, A. & Meyer, A. S. (in press) A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22(1). [WJML, rPFM]
Levin, J. (1985) A metrical theory of syllabicity. Unpublished Ph. D. dissertation, MIT. [AC-M]
Levitt, A. G. & Wang, Q. (1991) Evidence for language-specific rhythmic influences in the reduplicative babbling of French- and English-learning infants. *Language and Speech* 34:235–49. [LMe]
Liberman, A. M. & Mattingly, I. G. (1985) The motor theory of speech perception revised. *Cognition* 21:1–36. [aPFM, GR]
Lichtheim, L. (1885) On aphasia. *Brain* 7:433–84. [PL]
Lieberman, P. (1968) Primate vocalizations and human linguistic ability. *Journal of the Acoustical Society of America* 44:1157–64. [PL]
(1980) On the development of vowel production in young children. In: *Child phonology. Vol. 1: Production,* ed. G. H. Yeni-Komshian, J. F. Kavanagh & C. A. Ferguson. Academic Press. [JP]



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

(1984) *The biology and evolution of language.* Harvard University Press. [aPFM, GR]
(1994) Biologically bound behavior, free-will, and human evolution. In: *Conflict and cooperation in nature,* ed. J. I. Casti. Wiley. [PL]
(1995) What primate calls can tell us about human evolution. In: *Current topics in primate vocal communication,* ed. E. Zimmermann, J. D. Newman & U. Jürgens. Plenum Press. [PL]
(1998) *Eve spoke: Human language and human evolution.* Norton. [PL]
(in press) *The functional language system of the human brain: Its nature and evolution.* Harvard University Press. [PL]
Lieberman, P., Kako, E. T., Friedman, J., Tajchman, G., Feldman, L. S. & Jiminez, E. B. (1992) Speech production, syntax comprehension, and cognitive deficits in Parkinson's disease. *Brain and Language* 43:169–89. [PL]
Lim, S., Dinner, D., Pillay, P., Luders, H., Morris, H., Klein, G., Wylie, E. & Awad, I. (1994) Functional anatomy of the human supplementary sensorimotor area: Results of extraoperative electrical stimulation. *Electroencephalography and Clinical Neurophysiology* 91:179–93. [JHA]
Lin, L.-D., Murray, G. M. & Sessle, B. J. (1994) Functional-properties of single neurons in the primate face primary somatosensory cortex. 2. Relations with different directions of trained tongue protrusion. *Journal of Neurophysiology* 71:2391–400. [SFW]
(in press) Effects on primate mastication of reversible inactivation by cooling of the face primary somatosensory cortex. *Archives of Oral Biology.* [BJS]
Lindblom, B. (1992) Phonological units as adaptive emergents of lexical development. In: *Phonological development: Models, research, implications,* ed. C. A. Ferguson, L. Menn & C. Stoel-Gammon. York Press. [rPFM]
Lindblom, B., Lubker, J. & Gay, T. (1979) Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. *Journal of Phonetics* 7:147–61. [JP]
Lindblom, B. & Maddieson, I. (1988) Phonetic universals in consonant systems. In: *Language, speech and mind,* ed. L. M. Hyman & C. N. Li. Routledge. [aPFM]
Locke, J. (1983) *Phonological acquisition and change.* Academic Press. [aPFM]
Locke, J. L, Bekken, K. E., McMinn-Larson, L. & Wein, D. (1995) Emergent control of manual and vocal-motor activity in relation to the development of speech. *Brain and Language* 51:498–508. [GG]
Lu, M., Preston, J. & Strick, P. (1994) Interconnections between prefrontal cortex and premotor areas in the frontal lobe. *Journal of Comparative Neurology* 341:375–92. [JHA]
Luders, H. O. (1996) Functions of the supplementary motor area: Summary of a workshop. In: *Advances in neurology, vol. 70: Supplementary sensorimotor area,* ed. H. O. Luders. Lippincott-Raven. [rPFM]
Luders, H. O., Lesser, R., Dinner, D., Morris, H., Wylie, E. & Godoy, J. (1988) Localization of cortical function: New information from extraoperative monitoring of patients with epilepsy. *Epilepsia* 29(suppl. 2):S56-S65. [JHA]
Lund, J. P. (1991) Mastication and its control by the brainstem. *Critical Reviews in Oral Biology and Medicine* 2:33–64. [JPL]
Lund, J. P. & Enomoto, S. (1988) The generation of mastication by the central nervous system. In: *Neural control of rhythmic movements,* ed. A. Cohen, S. Rossignol & S. Grillner. Wiley. [aPFM]
Lund, J. P. & Lamarre, Y. (1974) Activity of neurons in the lower precentral cortex during voluntary and rhythmic jaw movements in the monkey. *Experimental Brain Research* 19:282–89. [JHA, JPL]
Lund, J. P., Sasamoto, K., Murakami, T. & Olsson, K. Å. (1984) Analysis of rhythmical movements produced by electrical stimulation of the motor-sensory cortex of rabbits. *Journal of Neurophysiology* 52:1014–29. [JPL]
Luppino, G., Matelli, M., Camarda, R., Gallese, V. & Rizzolatti, G. (1991) Multiple representations of body movements in mesial area 6 and the adjacent cingulate cortex; an intra-cortical microstimulation study. *Journal of Comparative Neurology* 311:463–82. [JHA]
Luppino, G., Matelli, M., Camarda, R. & Rizzolatti, G. (1993) Cortical connections of area F3 (SMA proper) and area F6 (Pre-SMA) in the macaque monkey. *Journal of Comparative Neurology* 338:114–40. [JHA, AAG]
Luschei, E. S. & Goldberg, L. J. (1981) Neural mechanisms of mandibular control: Mastication and voluntary biting. In: *Handbook of physiology: The nervous system, vol. 2.* American Physiological Society. [aPFM, BJS]
Luschei, E. S. & Goodwin, G. (1975) Role of monkey precentral cortex in control of voluntary jaw movements. *Journal of Neurophysiology* 38:146–57. [JHA]
Lynch, M. P., Oller, D. K., Steffens, M. L. & Buder, E. H. (1995) Phrasing in prelinguistic vocalizations. *Developmental Psychobiology* 28:3–25. [KB]
MacKay, D. G. (1970) Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia 8*:323–50. [rPFM]
Maclean, P. D. (1982) On the origin and progressive evolution of the triune brain. In: *Primate brain evolution: Methods and concepts,* ed. E. Armstrong and D. Falk. Plenum. [aPFM]
MacNeilage, P. F. (1964) Typing errors as clues to serial ordering mechanisms in language behavior. *Language and Speech* 7:144–59. [aPFM]
(1973) Central processes controlling speech production during sleep and waking. In: *The psychophysiology of thinking: Studies of covert processes,* ed. F. J. McGuigan & R. A. Schoonover. Academic Press. [aPFM]
(1982) Speech production mechanisms in aphasia. In: *Speech motor control,* ed. S. Grillner, B. Lindblom, J. Lubker & A. Persson. Pergamon. [aPFM]
(1985) Serial ordering errors in speech and typing. In: *Phonetic linguistics,* ed. V. A. Fromkin. Academic Press. [aPFM]
(1986) Bimanual coordination and the beginnings of speech. In: *Precursors to early speech,* ed. B. Lindblom & R. Zetterstrom. Stockton Press. [aPFM]
(1987a) The evolution of hemispheric specialization for manual function and language. In: *Higher brain functions: Explorations of the brain's emergent properties,* ed. S. P. Wise. Wiley. [aPFM, BW]
(1987b) Speech: Motor control. In: *Encyclopedia of neuroscience,* ed. G. A. Adelman. Birkhauser. [aPFM]
(1989) Grasping in modern primates: The evolutionary context. In: *Vision and action: The control of grasping,* ed. M. A. Goodale. Ablex. [aPFM]
(1990) The gesture as a unit in speech perception theories. In: *Modularity and the motor theory,* ed. I. G. Mattingly & M. G. Studdert-Kennedy. Erlbaum. [aPFM]
(1991a) Articulatory phonetics. In: *Oxford international encyclopedia of linguistics,* ed. W. Bright. Oxford University Press. [aPFM]
(1991b) The "postural origins" theory of primate neurobiological asymmetries. In: *Biological and behavioral determinants of language,* ed. N. A. Krasnegor, D. M. Rumbaugh, R. L. Schiefelbusch & M. G. Studdert-Kennedy. Erlbaum. [aPFM]
(1992) Evolution and lateralization of the two great primate action systems. In: *Language origin: A multidisciplinary approach,* ed. J. Wind, .B. Chiarelli, B. Bichakjian & A. Nocentini. Kluwer. [aPFM]
(1994) Prolegomena to a theory of the sound pattern of the first language. *Phonetica* 51:184–94. [arPFM]
(1998a) Evolution of the mechanism of language output: Comparative neurobiology of vocal and manual communication. In: *Approaches to the evolution of language: Social and cognitive bases,* ed. J. R. Hurford, C. Knight & M. G. Studdert-Kennedy. Cambridge University Press. [aPFM, BW]
(1998b) Towards a unified view of cerebral hemispheric specializations in vertebrates. In: *Comparative neuropsychology,* ed. A. D. Milner. Oxford University Press (in press). [aPFM]
MacNeilage, P. F. & Davis, B. L. (1990a) Acquisition of speech production: Frames then content. In: *Attention and performance XIII: Motor representation and control,* ed. M. Jeannerod. Erlbaum. [rPFM]
(1990b) Acquisition of speech production: The achievement of segmental independence. In: *Speech production and speech modeling,* ed. W. J. Hardcastle & A. Marchal. Kluwer. [rPFM]
(1993) Motor explanations of babbling and early speech patterns. In: *Developmental neurocognition: Speech and face processing in the first year of life,* ed. B. Boysson-Bardies, S. de Schonen, P. Jusczyk, P. F. MacNeilage & J. Morton. Kluwer. [rPFM]
(1996) From babbling to first words: Phonetic patterns. *Proceedings of the first ECA tutorial and research workshop on speech production modelling. Autrans, France,* pp. 155–57. [aPFM]
MacNeilage, P. F., Davis, B. L. & Matyear, C. L. (1997) Babbling and first words: Phonetic similarities and differences. *Speech Communication* 22:269–77. [rPFM]
MacNeilage, P. F., Studdert-Kennedy, M. G. & Lindblom, B. (1984) Functional precursors to language and its lateralization. *American Journal of Physiology R* (*Regulatory Integrative and Comparative Physiology* 15) 246:912–15. [aPFM]
(1985) Planning and production of speech: An overview. In: *Planning and production of speech in normally hearing and deaf people,* ed. J.. Lauter. ASHA Reports. [aPFM]
(1987) Primate handedness reconsidered. *Behavioral and Brain Sciences* 10:247–63. [aPFM]
Maddison, I. & Precoda, K. (1990) Syllable structure and phonetic models. *Phonology* 9:45–60. [rPFM]
Maeda, S. & Boë, L.-J. (1997) Modèlization de la croissance du conduit vocal. *La voyelle dans tous ses états.* Nantes. [CA]
Maki, S. (1990) An experimental approach to the postural origins theory of neurobiological asymmetries in primates. Unpublished Ph. D. dissertation, University of Texas at Austin. [aPFM]
Malmberg, B. (1960) *Kort lärobok i fonetik.* Gleerups. [BL]
Marder, E. & Calabrese, R. L. (1996) Principles of rhythmic motor pattern generation. *Physiological Reviews* 76:687–717. [AAG]
Marler, P. (1977) The structure of animal communication sounds. In: *Recognition of complex acoustic signals,* ed. T. H. Bullock. Dahlem Konferenzen. [aPFM]
Marsden, C., Deeke, L., Freund, H., Hallet, M., Passingham, R., Shibaski, H., Tanji, J. & Wiesendanger, M. (1996) The functions of the supplementary motor area. In: *Advances in neurology, vol. 70: Supplementary motor area,* ed. H. Luders. Lippincott-Raven. [JHA]


https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

Martin, N., Dell, G. S., Saffran, E. M. & Schwartz, M. F. (1994) Origins of paraphasias in deep dysphasia: Testing the consequences of a decay impairment to an interactive spreading activation model of lexical retrieval. *Brain and Language* 47:609–60. [GG]

Martin, R. E., Murray, G. M., Kemppainen, P., Masuda, Y. & Sessle, B. J. (1997) Functional properties of neurons in the primate tongue primary motor cortex during swallowing. *Journal of Neurophysiology* 78:1516–30. [BJS]

Martin, R. E., Murray, G. M. & Sessle, B. J. (1995) Cerebral cortical control of primate orofacial movements: Role of face motor cortex in trained and semi-automatic motor behaviors. In: *Alpha and gamma motor systems,* ed. A. Taylor, M. Gladden & R. Durbaba. Plenum Press. [AAG]

Martin, R. E. & Sessle, B. J. (1993) The role of the cerebral cortex in swallowing. *Dysphagia* 8:195–202. [BJS]

Masataka, N. & Bloom, K. (1994) Acoustic properties that determine adults' preferences for 3-month-old infant vocalizations. *Infant Behavior and Development* 17:461–64. [KB]

Masataka, N. & Fujita, K. (1989) Vocal learning of Japanese and rhesus monkeys. *Behaviour* 109:191–99. [RJA, rPFM]

Matelli, M., Camarda, R., Glickstein, M. & Rizzolatti, G. (1986) Afferent and efferent projections of the inferior area 6 in the macaque monkey. *Journal of Comparative Neurology* 251:281–98. [AAG]

Matelli, M., Luppino, G. & Rizzolatti, G. (1991) Architecture of superior and mesial area 6 and adjacent cingulate cortex. *Journal of Comparative Neurology* 311:445–62. [JHA]

Matsuzaka, Y., Aizawa, H. & Tanji, J. (1992) A motor area rostral to the supplementary motor areas (pre-supplementary area) in the monkey: Neuronal activity during a learned motor task. *Journal of Neurophysiology* 68:653–62. [JHA]

Mayr, E. (1982) *The growth of biological thought.* Bellknap. [aPFM]

McCaffrey, H. L., Davis, B. L. & MacNeilage, P. F. (1997) Babbling and words before and after early cochlear implant: A case study. Paper presented at the Annual Meeting of the American Speech, Language and Hearing Association, Boston MA., November, 1997. [rPFM]

McCune, L. & Vihman, M. M. (1987) Vocal motor schemes. *Papers and Reports in Child Language Development* 26:72–79. [LM]

(1997) A critical role for phonetic development in the transition to reference (submitted). [LM]

McCune, L., Vihman, M. M., Roug-Hellechius, L., Delery, D. B. & Gogate, L. (1996) Grunt communication in human infants. *Journal of Comparative Psychology* 110:27–37. [LM]

McFarland, D. H. & Lund, J. P. (1995) Modification of mastication and respiration during swallowing in the adult human. *Journal of Neurophysiology* 74:1509–17. [JPL]

Medicus, G. (1992) The inapplicability of the biogenetic rule to behavioral development. *Human Development* 35:1–8. [aPFM]

Meier, R. P., McGarvin, L., Zakia, R. A. E. & Willerman, R. (1997) Silent mandibular oscillations in vocal babbling. *Phonetica* 54:153–71. [KB, aPFM]

Meijers, L. M. & Eijkman, E. G. (1974) The motor system in simple reaction time experiments. *Acta Psychologica* 38:367–77. [SG]

Meringer, R. & Mayer, K. (1895) *Versprechen und Verlesen. Eine psychologisch-linguistische Studie.* Göschen'sche Verlagsbuchhandlung. New edition with an introductory essay by A. Cutler & D. A. Fay (1978). John Benjamins. [WJML]

Mesulam, M. M. (1990) Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Annals of Neurology* 28:597–613. [PL]

Mills, D. L., Coffey-Corina, S. & Neville, H. J. (1994) Variability in cerebral organization during primary language acquisition. In: *Human behavior and the developing brain,* ed. G. Dawson & K. W. Fischer. Guilford Press. [LM]

Mohr, J. P. (1976) Broca's area and Broca's aphasia. In: *Studies in neurolinguistics, vol. 1.,* ed. H. Whitaker & H. A. Whitaker. Academic Press. [rPFM]

Moore, B. R. (1992) Avian movement imitation and a new form of mimicry: Tracing the evolution of a complex form of learning. *Behaviour* 122:231–63. [IMP]

Morgane, P. J., Jacobs, M. S. & Galaburda, A. M. (1985) Conservative features of neocortical evolution in the dolphin brain. *Brain, Behavior and Evolution* 26:176–84. [GG]

Muakkassa, K. & Strick, P. (1979) Frontal lobe inputs to the primate motor cortex: Evidence for four somatotopically organized "pre-motor" areas. *Brain Research* 177:176–82. [JHA]

Muller, J. (1848) *The physiology of the senses, voice and muscular motion with the mental faculties,* trans. W. Baly. Walton and Maberly. [PL]

Müller-Preuss, P. (1983) Inhibition of auditory neurones during phonation: Evidence of feed-forward mechanisms in brain processes controlling audio-vocal behaviour. In: *Advances in vertebrate neuroethology,* ed. J. P. Ewart, R. R. Capranica & J. P. Ingle. NATO ASI Series. [CA]

Munhall, K. G. & Ostry, D. J. (1983) Mirror-image movements in typing. In: *Cognitive aspects of skilled typewriting,* ed. W. E. Cooper. Springer-Verlag. [KGM]

Murphy, K., Corfield, D. R., Fink, G. R., Wise, R. J. S., Guz, A. & Adams, L. (1997) Neural mechanisms associated with the control of speech in man. *NeuroImage* 5(4):S253. [CA, rPFM]

Murray, G. M. & Sessle, B. J. (1992a) Functional properties of single neurons in the face primary motor cortex of the primate. 1. Input and output features of tongue motor cortex. *Journal of Neurophysiology* 67:747–58. [SFW]

(1992b) Functional properties of single neurons in the face primary motor cortex of the primate. 2. Relations with trained orofacial motor behavior. *Journal of Neurophysiology* 67:759–74. [BJS]

Mushiake, H., Inase, M. & Tanji, J. (1991) Neuronal activity in the primate premotor, supplementary motor, and precentral motor cortex during visually-guided and internally determined sequential movements. *Journal of Neurophysiology* 66:705–18. [AAG]

Myers, R. E. (1976) Comparative neurology of vocalization and speech: Proof of a dichotomy. *Annals of the New York Academy of Sciences* 280:745–57. [aPFM]

Naeser, M. A., Alexander, M. P., Helms-Estabrooks, N., Levine, H. L., Laughlin, S. A. & Geschwind, N. (1982) Aphasia with predominantly subcortical lesion sites; description of three capsular/putaminal aphasia syndromes. *Archives of Neurology* 39:2–14. [PL]

Nakamura, Y. & Katakura, N. (1995) Generation of masticatory rhythm in the brainstem. *Neuroscience Research* 23:1–20. [UJ]

Napier, J. R. (1962) The evolution of the hand. *Scientific American* 207:56–62. [aPFM]

Natsopoulos, D., Grouios, G., Bostantzopoulou, S., Mentenopoulos, G., Katsarou, Z. & Logothetis, J. (1993) Algorithmic and heuristic strategies in comprehension of complement clauses by patients with Parkinson's disease. *Neuropsychologia* 31:951–64. [PL]

Negus, V. E. (1949) *The comparative anatomy and physiology of the larynx.* Hafner. [aPFM]

Nespoulous, J-L., Lecours, A. R., Lafond, D. & Joanette, Y. (1985) Jargonagraphia with(out) jargonaphasia. Paper presented at the BABBLE conference, Niagara Falls, March 1985. [aPFM]

Netsell, R. (1981) The acquisition of speech motor control: A perspective with directions for research. In: *Language behavior in infancy and early childhood,* ed. R. E. Stark. Elsevier-North Holland. [KB]

Nicolelis, M. A. L., Baccala, L. A., Lin, R. C. S. & Chapin, J. K. (1995) Sensorimotor encoding by synchronous neural ensemble activity at multiple levels of the somatosensory system. *Science* 268:1353–58. [AAG]

Nicolelis, M. A. L., Ghazanfar, A. A., Faggin, B. M., Votaw, S. & Oliveira, L. M. O. (1997) Reconstructing the engram: Simultaneous, multisite, many single neuron recordings. *Neuron 18*:529–37. [AAG]

Nieuwenhuys, R. (1972) Topographical analysis of the brain stem of the lamprey *Lampetra fluviulis. Journal of Comparative Neurology* 145:165–78. [JPL]

Nittrouer, S., Studdert-Kennedy, M. G. & McGowan, R. S. (1989) The emergence of phonetic segments. Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research* 32:120–32. [aPFM]

Norman, D. A. & Rumelhart, D. E. (1983) Studies of typing from the LNR research group. In: *Cognitive aspects of skilled typewriting,* ed. W. E. Cooper. Springer-Verlag. [aPFM]

Nottebohm, F. (1980) Brain pathways for vocal learning in birds: A review of the first ten years. *Progress in Psychobiology and Physiological Psychology* 9:85–124. [IMP]

Ohala, J. J. (1978) Phonological notations as models. In: *Proceedings of the Twelfth International Congress of Linguists, Innsbruck,* ed. W. U. Dressler & W. Meid. Institut für Sprachwissenschaft der Universitat Innsbruck. [aPFM]

Ohala, J. J. (1995) Speech perception is hearing sounds, not tongues. *Journal of the Acoustical Society of America* 99:1718–25. [JJO]

Ohala, J. J. & Kawasaki-Fukumori, H. (1997) Alternatives to the sonority hierarchy for explaining segmental sequential constraints. In: *Language and its ecology: Essays in memory of Einar Haugen,* ed. S. Eliasson & E. H. Jahr. Mouton de Gruyter. [JJO]

Ojemann, G. (1983) Brain organization for language from the perspective of electrical stimulation mapping. *Behavioral and Brain Sciences* 6:189–230. [aPFM]

Oller, D. K. & Eilers, R. E. (1988) The role of audition in infant babbling. *Child Development* 59:441–49. [LMe]

Orgogozo, J. M. & Larsen, B. (1979) Activation of the supplementary motor area during voluntary movement in man suggests it works as a supramotor area. *Science* 206:847–50. [aPFM]

Owren, M. J., Dieter, J. A., Seyfarth, R. M. & Cheney, D. L. (1992) "Food" calls produced by adult female rhesus (*Macaca mulatta)* and Japanese (*M. fuscata)* macaques, their normally reared offspring, and offspring cross-fostered between species. *Behaviour* 120:218–31. [rPFM]

(1993) Vocalizations of rhesus (*Macaca mulatta)* and Japanese (*M. fuscata)*



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

macaques cross-fostered between species show evidence of only limited modification. *Developmental Psychobiology* 26:389–406. [rPFM]
Pandya, D. (1987) Association cortex. In: *The encyclopedia of neuroscience,* ed. G. Adelman. Birkhauser [aPFM]
Pardo, J. & Fox, P. (1993) Preoperative assessment of the cerebral hemisphere dominance for language with CBF PET. *Human Brain Mapping* 1:57–68. [JHA]
Passingham, R. E. (1987) Two cortical systems for directing movement. In: *Ciba Foundation Symposium No. 132,* ed. G. Bock, M. O'Connor & J. Marsh. Wiley. [aPFM]
Pastor, M. A., Artieda, J., Jahanshahi, M. & Obeso, J. A. (1992) Time estimation and reproduction is abnormal in Parkinson's disease. *Brain 1*15:211–25. [GG]
Patterson, D. K. & Pepperberg, I. M. (1998, in press) Acoustic and articulatory correlates of stop consonants in a parrot and a human subject. *Journal of the Acoustical Society of America* 103. [IMP]
Patterson, D. K., Pepperberg, I. M., Story, B. H. & Hoffman, E. (1997) How parrots talk: Insights based on CT scans, image processing, and mathematical models. In: *SPIE Proceedings, vol. 3033,* ed. E. Hoffman. International Society for Optical Engineering. [IMP]
Paulesu, E., Frith, C. D. & Frackowiak, R. S. J. (1993) The neural correlates of the component of working memory. *Nature* 362:342–44. [aPFM]
Paus, T., Petrides, M., Evans, A. C. & Meyer, E. (1993) The role of the human anterior cingulate cortex in the control of oculomotor, manual and speech responses: A positron emission tomography study. *Journal of Neurophysiology* 70:453–69. [aPFM]
Pearson, K. (1987) Central pattern generation: A concept under scrutiny. In: *Advances in physiological research,* ed. H. McLennen. Plenum Press. [AAG]
Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V. & Rizzolatti, G. (1992) Understanding motor events: A neurophysiological study. *Experimental Brain Research* 91:176–80. [aPFM]
Penfield, W. & Jasper, H. (1954) *Epilepsy and the functional anatomy of the human brain.* Little Brown. [aPFM]
Penfield, W. & Roberts, L. (1959) *Speech and brain mechanisms.* Princeton University Press. [arPFM]
Penfield, W. & Welch, K. (1951) The supplementary motor area of the cerebral cortex: A clinical and experimental study. *A. M. A. Archives of Neurology and Psychiatry* 66:289–317. [aPFM]
Pepperberg, I. M. (1990) Some cognitive capacities of an African Grey parrot. In: *Advances in the study of behavior, vol. 19,* ed. P. J. B. Slater, J. S. Rosenblatt & C. Beer. Academic Press. [IMP]
Pepperberg, I, M., Brese, K. J. & Harris, B. J. (1991) Solitary sound play during acquisition of English vocalizations by an African Grey parrot (*Psittacus erithacus*): Possible parallels with children's monologue speech. *Applied Psycholinguistics* 12:151–78. [IMP]
Peters, J. (in preparation) Acoustic determinants of the perception of infant vowels. [JP]
Petersen, M., Beeche, M., Zoloth, S., Moody, D. & Stebbins, W. (1978) Neural lateralization of species-specific vocalizations by Japanese macaques (*Macaca fuscata*). *Science* 202:324–26. [JHA]
Petersen, S. E., Fox, P. T., Posner, M. I., Mintum, M. & Raichle, M. E. (1988) Positron emission tomographic studies of the cortical anatomy of single word processing. *Nature* 331:585–89. [aPFM]
(1989) Positron emission tomographic studies of the processing of single words. *Journal of Cognitive Neuroscience* 1(2):153–70. [JHA]
Petitto, L. & Marentette, P. (1991) Babbling in the manual mode: Evidence for the ontogeny of language. *Science* 251:1493–96. [aPFM]
Petropoulos, D., Lund, J. P. & Dubuc, R. (1997) The neural correlate of fictive feeding in lampreys. *Society for Neuroscience Abstracts* 23:206. [JPL]
Picard, N. & Strick, P. (1997) Activation of the medial wall during remembered sequences of reaching movements in monkeys. *Journal of Neurophysiology* 77:2197–2201. [JHA, JPL]
Pickett, E. R., Kuniholm, Protopapas, A., Friedman, J. & Lieberman, P. (1998) Selective speech motor, syntax and cognitive deficits associated with bilateral damage to the head of the caudate nucleus and the putamen. A single case study. *Neuropsychologia* 36:173–88 [PL]
Pimiäa, P. (1990) Semantic features of some mouth patterns in Finnish Sign Language. In: *Current trends in European sign language research,* ed. S. Prillwitz & T. Vollhaber. Signum Press. [BW]
Pinker, S. & Bloom, P. (1990) Natural language and natural selection. *Behavioral and Brain Sciences* 13:707–84. [PL, aPFM]
Ploog, D. (1992) The evolution of vocal communication. In: *Nonverbal vocal communication: Comparative and developmental approaches,* ed. M. Papousek, H. Papousek and U. Jürgens. Cambridge University Press. [CA]
Poeck, K., de Bleser, R. & Graf von Keyserlingk, D. (1984) Neurolinguistic status and localization of lesion in aphasic patients with exclusively consonant-vowel recurring utterances. *Brain 1*07:199–217. [CA, rPFM]
Poizner, H., Klima, E. & Bellugi, U. (1987) *What the hands reveal about the brain.* MIT Press. [aPFM]
Porter, R. J. & Castellanos, F. X. (1980) Speech production measures of speech perception: Rapid shadowing of VCV syllables. *Journal of the Acoustical Society of America* 67:1349–56. [aPFM]
Posner, M. I., Petersen, S. E., Fox, P. T. & Raichle, M. E. (1988) Localization of cognitive operations in the human brain. *Science* 240:1627–31. [aPFM]
Premack, D. (1976) Language and intelligence in ape and man. *American Scientist* 64:674–93. [JHA]
(1986) *Gavagai!* MIT Press. [aPFM]
Price, C., Wise, R., Howard, D., Patterson, K., Watson, K. & Frackowiak, R. S. J. (1993) The brain regions involved in the recognition of visually presented words. *Journal of Cerebral Blood Flow and Mechanism* 13(suppl. 1):S501. [aPFM]
Redican, W. K. (1975) Facial expressions in nonhuman primates. In: *Primate behavior: Developments in field and laboratory research, vol. 4,* ed. L. A. Rosenblum. Academic Press. [aPFM]
Regan, D. & Tyler, C. W. (1971) Temporal summation and its limit for wavelength changes: An analog of Bloch's law for color vision. *Journal of the Optical Society of America* 61:1414–21. [SG]
Rizzolatti, G. & Arbib, M. A. (1998) Language in a grip. *Trends in Neurosciences* (in press). [GR]
Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G. & Matelli, M. (1988) Functional organization of inferior area 6 in the macaque monkey II. Area F5 and the control of distal movements. *Experimental Brain Research* 71:491–507. [aPFM]
Rizzolatti, G., Fadiga, L., Gallese, V. & Fogassi, L. (1996a) Premotor cortex and the recognition of motor action. *Cognitive Brain Research* 3:131–41. [CA]
Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D. & Fazio, F. (1996b) Localization of grasp representation in humans by positron emission tomography: 1. Observation versus execution. *Experimental Brain Research* 111:246–52. [GR]
Rizzolatti, G., Luppino, G. & Matelli, M. (1996c) The classic supplementary motor area 1 is formed by two independent areas. In: *Advances in neurology, vol. 70: Supplementary motor area,* ed. H. Luders. Lippincott-Raven. [JHA]
Rizzolatti, G., Matelli, M. & Pavesi, G. (1983) Deficits in attention and movement following the removal of postarcuate (area 6) and prearcuate (area 8) cortex in macaque monkeys. *Brain* 106:655–73. [aPFM]
Robinson, B. W. (1976) Limbic influences on human speech. *Annals of the New York Academy of Sciences* 280:761–71. [aPFM]
Robinson, J. G. (1979) An analysis of the organization of vocal communication in the titi monkey (*Callicebus moloch*). *Zeitschrift fur Tierpsychologie* 49:381–405. [aPFM]
Roelofs, A. & Meyer, A. S. (in press) Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* [WJML]
Rohrbaugh, J. W., Donchin, E. & Eriksen, C. W. (1974) Decision making and the P300 component of the cortical evoked response. *Perception and Psychophysics* 15:368–74. [SG]
Roland, P. (1993) *Brain activation.* Wiley-Liss. [aPFM]
Roland, P. E., Larsen, B., Lassen, N. A. & Skinhoj, E. (1980) Supplementary motor area and other cortical areas in organization of voluntary movements in man. *Journal of Neurophysiology* 43:118–36. [UJ, rPFM]
Ross, E. D. (1993) Nonverbal aspects of language. *Neurological Clinics* 11:9–23. [JSS]
Rossignol, S., Lund, J. P. & Drew, T. (1988) The role of sensory inputs in regulating patterns of rhythmical movements in higher vertebrates: A comparison between locomotion, respiration and mastication. In: *Neural control of rhythmic movements in vertebrates,* ed. A. Cohen, S. Rossignol & S. Grillner. Wiley. [aPFM]
Rubens, A. B. (1975) Aphasia with infarction in the territory of the anterior cerebral artery. *Cortex* 11:239–50. [UJ]
Rumelhart, D. E. & Norman, D. A. (1982) Simulating a skilled typist: A study of skilled cognitive-motor performance. *Cognitive Science* 6:1–36. [aPFM]
Russell, B. (1905) On denoting. *Mind 1*4:479–93. [AC-M]
Ryle, G. (1960) Letters and syllables in Plato. *Philosophical Review* 69:431–51. [AC-M]
Scheibel, A. B. (1993) Dendritic structure and language development. In: *NATO ASI Series: D. Developmental neurocognition: Speech and face processing in the first year of life,* ed. B. de Boysson-Bardies, S. de Schonen, Jusczyk, P. F. MacNeilage & J. Morton. Kluwer. [LM]
Schermer, T. (1990) *In search of a language: Influences from spoken Dutch on Sign Language of the Netherlands.* Eburon. [BW]
Schiller, N. O., Meyer, A. S., Baayen, R. H. & Levelt, W. J. M. (1996) A comparison of lexeme and speech syllables in Dutch. *Journal of Quantitative Linguistics* 3:8–28. [WJML]
Schiller, N. O., Meyer, A. S. & Levelt, W. J. M. (1997) The syllabic structure of


https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

spoken words: Evidence from the syllabification of intervocalic consonants. *Language and Speech* 40:103–40. [WJML]
Schott, D. (1975) Quantitative analysis of the vocal repertoire of squirrel monkeys (*Saimiri sciureus*). *Zeitschrift fur Tierpsychologie* 38:225–50. [rPFM]
Schreiner, C. E. & Urbas, J. V. (1988) Representation of amplitude modulation in the auditory cortex of the cat. I. The anterior auditory field (AAF). *Hearing Research* 21:227–41. [SG]
Schwartz, M F., Saffran, E. M., Bloch, D. E. & Dell, G. S. (1994) Disordered speech production in aphasic and normal speakers. *Brain and Language* 47:52–88. [GG]
Schwarz, G. & Lund, J. P. (1995) Modification of rhythmical jaw movements by noxious pressure applied to the zygoma in decerebrate rabbits. *Pain* 63:151–61. [JPL]
Searleman, A. (1980) Subject variables and cerebral organization for language. *Cortex* 16:239–54. [aPFM]
Sergent, J., Zuck, E., Levesque, M. & Macdonald, B. (1992) Positron emission tomography study of letter and object processing: Empirical findings and methodological considerations. *Cerebral Cortex* 2:68–80. [aPFM]
Shallice, T. (1988) *From neuropsychology to mental structure.* Cambridge University Press. [aPFM]
Shattuck-Hufnagel, S. (1979) Speech errors as evidence for a serial ordering mechanism in sentence production. In: *Sentence processing: Psycholinguistic studies presented to Merrill Garrett,* ed. W. E. Cooper & E. C. T. Walker. Erlbaum. [WJML, aPFM]
(1980) Serial ordering limits for speech production: Segments or other syllabic subunits. *Journal of the Acoustical Society of America* 68(suppl. 1):S32. [aPFM]
Shattuck-Hufnagel, S. & Klatt, D. H. (1979) The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior* 18:41–55. [aPFM]
Sieratzki, J. S. & Woll, B. (1996) Why do mothers cradle babies on their left? *The Lancet* 347(9017):746–48. [JSS]
(1998, in press) Toddling into language. Precocious language development in children with Spinal Muscular Atrophy. In: *Proceedings of the 22nd Annual Boston University Conference on Language Development, vol. 2.* Cascadilla Press. [JSS]
Smith, C., Browman, C., McGowan, R. & Kay, B. (1993) Extracting dynamic parameters from speech movement data. *Journal of the Acoustical Society of America* 93:1580–88. [SG]
Snowdon, C., Brown, C. & Petersen, M. (1982) *Primate communication.* Cambridge University Press. [JHA]
Snyder, P. J. & Harris, L. J. (1997) Lexicon size and its relation to foot preference in the African Grey parrot (*Psittacus erithacus*). *Neuropsychologia* 35:919–26. [IMP]
Sokoloff, A. J. (1991) Musculotopic organization of the hypoglossal nucleus in the grass frog, *rana-pipiens. Journal of Comparative Neurology* 308:505–12. [SFW]
Sokoloff, A. J. & Deacon, T. W. (1992) Musculotopic organization of the hypoglossal nucleus in the cynomolgus monkey, *macaca-fascicularis. Journal of Comparative Neurology* 324:81–93. [SFW]
Stark, R. E. (1980) Stages of speech development in the first year of life. In: *Phonological development: I. Production,* ed. G. Yeni-Komshian, J. Kavanagh & C. A. Ferguson. Academic Press. [LMe]
Stengel, E. & Lodge-Patch, I. C. (1955) "Central" aphasia associated with parietal symptoms. *Brain* 78:401–16. [aPFM]
Stetson, R. H. (1951/1988) *Motor phonetics.* College Hill Press. [KGM]
Stevens, K. N. (1989) On the quantal nature of speech. *Journal of Phonetics* 17:3–46. [aPFM]
Stoel-Gammon, C. & Otomo, K. (1986) Babbling development of hearing-impaired and normally-hearing subjects. *Journal of Speech and Hearing Disorders* 51:33–41. [LMe]
Strawson, P. F. (1950) On referring. *Mind* 54:320–44. [AC-M]
Striedter, G. (1994) The vocal control pathways in budgerigars differ from those in songbirds. *Journal of Comparative Neurology* 343:35–56. [IMP]
Stuss, D. T. & Benson, D. F. (1986) *The frontal lobes.* Raven Press. [PL, aPFM]
Sutton, D., Larson, C. & Lindeman, R. (1974) Neocortical and limbic lesion effects on primate phonation. *Brain Research* 16:61–75. [JHA, CA]
Talairach, J. & Tournoux, P. (1988) *Co-planar stereotaxic atlas of the human brain. 3 dimensional system. An approach to cerebral imaging.* Translated by Mark Rayport. Thieme Medical Publishers Inc. [aPFM]
Tanji, J. (1994) The supplementary motor area in the cerebral cortex. *Neuroscience Research* 19:251–68. [JHA]
Tanji, J., Shima, Y., Matsuzaka, Y. & Halsband, U. (1995) Neuronal activity in the supplementary, presupplementary, and premotor cortex of monkey. In: *Functions of the cortico-basal ganglia loop,* ed. M. Kimura & A. M. Graybiel. Springer. [aPFM]
Taylor, J. R. (1996) *Possessives in English: An exploration in cognitive grammar.* Clarendon Press. [AC-M]
Thelen, E. (1981) Rhythmical behavior in infancy: An ethological perspective. *Developmental Psychology* 17:237–57. [aPFM, LM]
(1991) Motor aspects of emergent speech: A dynamic approach. In: *Biobehavioral foundations of language,* ed. N. Krasnegor. Erlbaum. [AAG]
Tobias, P. V. (1987) The brain of *homo habilis:* A new level of organization in cerebral evolution. *Journal of Human Evolution* 16:741–61. [aPFM]
Tulving, E. (1995) Organization of memory: Quo vadis? In: *The cognitive neurosciences,* ed. M. S. Gazzaniga. MIT Press. [JSS]
Urban, P. P., Hopf, H. C., Connemann, B., Hundemer, H. P. & Koehler, J. (1996) The course of cortico-hypoglossal projections in the human brain-stem: Functional testing using trancranial magnetic stimulation. *Brain* 119:1031–38. [SFW]
Vihman, M. (1992) Early syllables and the construction of phonology. In: *Phonological development: Models, research, implications,* ed. C. A. Ferguson, L. Menn & C. Stoel-Gammon. York Press. [LM]
Vihman, M. M. & Velleman, S. (1989) Phonological reorganization: A case study. *Language and Speech* 32:149–70. [LM]
Vihman, M. M. (1993) Vocal motor schemes, variation, and the production-perception link. *Journal of Phonetics* 21:163–69. [LMe]
Vihman, M. M., Velleman, S. & McCune, L. (1994) How abstract is child psychology? In: *First and second language phonology,* ed. M. Yavas. Singular Publishing Group. [LM]
Vogt-Svendsen, M. (1983) Lip movements in Norwegian Sign Language. In: *Language in sign,* ed. J. G. Kyle & B. Woll. Croom Helm. [BW]
Vousden, J. L., Brown, G. D. A. & Harley, T. A. (1997) Serial control of phonology in speech production: A hierarchical model. Manuscript to be submitted for publication. [TAH]
Walker, A. & Green, H. (1938) Electrical excitability of the motor face area: A comparative study in primates. *Journal of Neurophysiology* 1:152–65. [JHA]
Walker, S. F. (1994) Animal communication. In: *The encyclopedia of language and linguistics, vol. 4,* ed. R. E. Asher & J. M. Y. Simpson. Pergamon. [SFW]
Warren, D. K., Patterson, D. K. & Pepperberg, I. M. (1996) Mechanisms of American English vowel production in a Grey parrot (*Psittacus erithacus*). *Auk* 113:41–58. [IMP]
Watson, C. (1975) The role of the precentral gyrus in the control of facial movements in *Macaca mulatta.* Unpublished Ph. D. dissertation, University of Chicago. [JHA]
Watson, R. T., Fleet, W. S., Gonzales-Rothi, L. & Heilman, K. M. (1986) Apraxia and the supplementary motor area. *Archives of Neurology* 43:787–92. [aPFM]
Weiss, D. A. (1951) Chewing and the origin of speech. In: *The chewing approach in speech and voice therapy,* ed. D. A. Weiss & H. H. Beebe. Karger. [JJO]
West, R. A. & Larson, C. R. (1995) Neurons of the anterior mesial cortex related to faciovocal activity in the awake monkey. *Journal of Neurophysiology* 74:1856–69. [SFW]
Westberg, K-G., Olsson, K. Å., Lund, J. P. & Clavelou, P. (1995) Premotor neurons in the oral nucleus of the spinal trigeminal tract: Functional characteristics. In: *Brain and oral functions,* ed. T. Morimoto, T. Matsuya & T. Takada. Elsevier. [JPL]
White, S. S. (1968a) Movements of the larynx during crowing in the domestic cock. *Journal of Anatomy* 103:390–92. [JS]
(1968b) Mechanisms involved in deglutition in *Gallus domesticus. Journal of Anatomy* 104:177. [JS]
White, S. S. & Chubb, J. C. (1968) The muscles and movements of the larynx of Gallus domesticus. *Journal of Anatomy* 102:575. [JS]
Wilkins, W. K. & Wakefield, J. (1995) Brain evolution and neurolinguistic preconditions. *Behavioral and Brain Sciences* 18:161–226. [SFW]
Winter, P., Ploog, D. & Latta, J. (1966) Vocal repertoire of the squirrel monkey (*Saimiri sciureus*), its analysis and significance. *Experimental Brain Research* 1:359:84. [rPFM]
Wise, R. J., Chollet, F., Hadar, U., Friston, K., Hoffner, E. & Frackowiak, R. (1991) Distribution of cortical neural networks involved in word comprehension and word retrieval. *Brain* 114:1803–17. [aPFM]
Wise, S., Fried, I., Oliver, A., Paus, T., Rizzolatti, G. & Zilles, K. (1996) Workshop on the anatomic definition and boundaries of the supplementary sensorimotor area. In: *Advances in neurology, vol. 70: Supplementary motor area,* ed. H. Luders. Lippincott-Raven. [JHA]
Wolff, P. H. (1967) The role of biological rhythms in psychological development. *Bulletin of the Menninger Clinic* 31:197–218. [aPFM]
(1968) Stereotypic behavior and development. *Canadian Psychologist* 9:474–83. [aPFM]
Woll, B. (1993) The oral component in sign languages: Sonority of mouth and hand. Paper presented at the Amsterdam Workshop on the Phonology and Morphology of Sign Language. [BW]
(1996) Do sign languages tell us anything about the origins of human language? Paper presented at the Evolution of Human Language Conference, University of Edinburgh, April 1996. [BW]
Woll, B. & Kyle, J. G. (1989) Communication and language development in children of deaf parents. In: *The social and cognitive aspects of normal and*



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press

*atypical language development*, ed. S. von Tetzchner, L. S. Siegel & L. Smith. Springer. [JSS]

Woolsey, C. N. (1958) Organization of somatic sensory and motor areas of the cerebral cortex. In: *Biological and biochemical bases of behavior*, ed. H. F. Harlow & C. N. Woolsey. University of Wisconsin Press. [aPFM]

Woolsey, C. N., Erickson, T. C. & Gilson, W. E. (1979) Localization in somatic sensory and motor areas of human cerebral cortex as determined by direct recording of evoked potentials and electrical stimulation. *Journal of Neurosurgery* 51:476–506. [aPFM]

Yeni-Komshian, G. & Benson, D. (1976) Anatomical study of cerebral asymmetry in the temporal lobe of humans, chimpanzees and rhesus monkeys. *Science* 192:387–89. JHA]

Zatorre, R. J., Evans, A. C., Meyer, E. & Gjedde, A. (1992) Lateralization of phonetic and pitch discrimination in speech processing. *Science* 256:846–49. [aPFM]

Zhang, S. P., Bandler, R. & Davis, P. J. (1995) Brain-stem integration of vocalization: Role of the nucleus retroambigualis. *Journal of Neurophysiology* 74:2500–12. [SFW]

Ziegler, W., Kilian, B. & Deger, K. (1997) The role of left mesial frontal cortex in fluent speech: Evidence from a case of left supplementary motor area hemorrhage. *Neuropsychologia* 35:1197–208. [rPFM]

Zilles, K., Schlaug, G., Geyer, S., Luppino, G., Matelli, M., Qu, M., Schleicher, A. & Schormann, T. (1996) Anatomy and transmitter receptors of the supplementary motor areas in the human and nonhuman primate brain. In: *Advances in neurology, vol. 70: Supplementary motor area*, ed. H. Luders. Lippincott-Raven. [JHA]

Zlatic, L., MacNeilage, P. F., Matyear, C. & Davis, B. L. (1997) Babbling of twins in a bilingual environment. *Applied Psycholinguistics* 18:453–69. [aPFM]

Zweers, G. A. (1982a) Drinking in the pigeon (*Columba livia* L. ). *Behaviour* 80:274–317. [JS]

(1982b) Pecking of the pigeon (*Columba livia* L. ). *Behaviour* 81:173–230. [JS]

Zweers, G. A., van Pelt, H. C. & Beckers, A (1981) Morphology and mechanics of the larynx of the pigeon (*Columba livia* L.): A drill-chuck system. *Zoomorphology* 99:37–69. [JS]



https://doi.org/10.1017/S0140525X98221261 Published online by Cambridge University Press