On the Evolution of Human Speech: Its Anatomical and Neural Bases

by Philip Lieberman

Fully human speech involves species-specific anatomy deriving from the descent of the tongue into the pharynx. The human tongue’s shape and position yields the 1:1 oral-to-pharyngeal proportions of our supralaryngeal vocal tract (SVT). Speech also necessitates a brain that can "reiterate," i. e. freely reorder a finite set of motor gestures to form a potentially infinite number of words and sentences. The end points of the evolutionary process are clear. Chimpanzees lack a SVT that can produce "quantal" sounds which facilitate both speech production and perception, and brains that can reiterate the phonetic contrasts apparent in their fixed vocalizations. The traditional Broca-Wernicke brain-language theory is incorrect; neural circuits linking regions of cortex with the basal ganglia and other subcortical structures regulate motor control, including speech production, as well as cognitive processes including syntax. The dating of the human form of the FOXP2 gene, which governs the embryonic development of these subcortical structures, provides an insight on the evolutionary history of speech and language. Speech most likely has a long evolutionary history. The starting points for human speech and, language perhaps were walking and running. However, fully human speech anatomy first appears in the fossil record in the Upper Paleolitic (about 50,000 years ago) and was absent in both Neanderthals and earlier humans.

PHILIP LIEBERMAN is the Fred M. Seed Professor of Cognitive and Linguistic Sciences and a Professor of Anthropology at Brown University (Providence RI 029012-1978, U. S. A. [Philip_Lieberman@brown.edu]. Born in 1934 he was educated at The Massachusetts Institute of Technology (BSEE and MSEE, 1958, Ph.D Linguistics, 1966). His publications include *The Biology and Evolution of Language* (Harvard University Press, 1984), *Human Language and Our Reptilian Brain: The subcortical Bases Speech, Syntax and Thought* (Harvard University Press, 2000), and *Toward an evolutionary biology of language* (Harvard University Press, 2006). His recent research, noted in this paper, has been supported by NASA under grant NCC9-58 with the National Space Biomedical Research Institute.

Although the focus of current linguistic research is syntax, speech is the derived feature of language, absent in even closely related living species. Speech allows us to transmit information at a rate faster than that which otherwise could otherwise be vocally transmitted. Speech also keeps words active in the neural computational space, "verbal working memory," in which the meaning of a sentence is discerned (Baddeley, 1986).The neural substrate that regulates speech production appears to play a part in syntactic operations and other cognitive processes. Therefore, any account of the evolution of human language must account for the specialized anatomy and neural mechanisms that make speech possible.

I will briefly review the anatomy and physiology of speech, focusing on the species-specific anatomy of the human "supralaryngeal vocal tract" (SVT). I will then discuss reconstructing the SVTs of fossil hominids, taking account of recent studies of human ontogenetic development and the constraints imposed by swallowing. The findings of these studies

provide a quantitative basis for inferring the speech producing anatomy of Neanderthals and other fossil hominids

I will then discuss the neural substrate that regulates speech production. Current findings refute the traditional theory localizing the neural bases of human language to Broca's and Wernickes areas. These areas of the cortex play a role in speech and language. However, they work in concert with other neural structures in circuits that link activity in these and other cortical areas to the basal ganglia and other subcortical structures. Evidence from neurophysiologic and behavioral studies of humans and other species show that the basal ganglia confer the "reiterative" quality of human speech, allowing us to reorder a finite number of learned motor acts to form an almost unbounded store of words. Chomsky's most recent candidate for the productive capacity of syntax is a "narrow faculty of language" (FNL), specific to humans and to syntax (Hauser, Chomsky and Fitch, 2002).The proposal here is that cortical-striatal-cortical neural circuits regulate syntax as well as speech production, yielding the productive qualities of syntax. Similar neural circuits grant cognitive flexibility and make possible seemingly unrelated human capacities such as composing music or dancing. I will endeavor to show that the evolutionary root of these human qualities is motor control. In this, I claim no original insight; the credit goes to Karl Lashly, who in 1951 proposed that neural mechanisms originally adapted for motor control are the basis for syntax and human creative behavior. The isolation and dating of the human form of the FOXP2 gene, which governs the embryonic development of the subcortical structures that support these neural circuits, provides insights on the evolution of human speech, language and cognition.

THE PHYSIOLOGY OF SPEECH

The vocal signals of all terrestrial mammals are generated by filtering a "source" of acoustic energy through an airway through which maximum energy passes at frequencies termed “formants” (Fant, 1960). For phonated sounds the source is a quasi-periodic series of "puffs" of air generated by rapidly opening and closing the vocal "folds" or "cords" of the larynx. The average fundamental frequency of phonation (F0), the rate at which these puffs of air occur, is perceived as the pitch of a person's voice. In many languages, such as those of China, words are differentiated by changes in F0 over the course of a syllable, but vowel quality is largely conveyed by formant frequency patterns, enhanced by distinctions in duration (Hellwag, 1781; Chiba and Kajiyama, 1941; Fant, 1960).

In humans, the airway above the larynx , the supralaryngeal vocal tract (SVT), continually changes its shape as we talk, producing a time-varying formant frequency pattern. Aperiodic noise generated at a constriction along the SVT can also serve as a source of acoustic energy that is filtered by the SVT; the sound transcribed by the phonetic symbol [h] in English is essentially a vowel having a noise source generated by air moving through a fixed laryngeal opening.

Formant Frequencies

In short, the larynx provides the source of acoustic energy for vowels and other phonated speech sounds; the SVT act as an acoustic filter that determines the phonetic quality of the sounds. A given SVT shape will let more acoustic energy through at a set of particular formant frequencies, local energy maxima, occurring in inharmonic combinations. The lowest formant frequency is

identified by the notation F1, the next highest as F2, the third as F3. For example, the vowels [i] and [u] of the words "see" and "sue", can be produced with identical F0's, -- different formant frequencies specify these vowels. As we talk,we change the SVTs shape and the resulting formant frequency pattern.

The relationship between formants, the laryngeal source and speech signals may be clearer if you think about how sunglasses work. The difference between a pair of sunglasses that make everything look blue and ones that make the world look pink is the balance of light energy frequencies that passes through the glasses. The tinted glass achieves these effects by "attenuating," i. e., reducing the amount of light energy throughout a range of frequencies. The combination of frequencies that are least attenuated determines the color. The same "source" of light, sunlight will provide a blue or pink world when filtered by different sunglasses. Think of formant frequencies as the acoustic frequencies that the SVT allows to pass through it with minimum attenuation.

The Supralaryngeal Vocal Tract (SVT)

The range of area functions and the overall length of the SVT determine the formant frequencies that it can generate. In the 18th and 19th centuries tubes were used to model the SVT. The tubes acted as acoustic filters; reeds as sources of acoustic energy. Computer-implemented models are now used to determine the formant frequencies that particular SVT shapes can produce (e. g., Henke, 1966; Stevens, 1972; Baer et al., 1991; Story, Titze, and Hoffman, 1996). The adult-like human SVT has a tongue having an almost circular sagital (midline) contour forming two segments, a "horizontal" oral cavity (SVTh), and a vertical pharyngeal cavity (SVTv) having almost equal length (1:1 proportions), positioned at

a right angle. Movements of the undistorted tongue in the space defined by the oral cavity and pharynx can produce the abrupt midpoint 10:1 area function discontinuities necessary to produce the formant frequency patterns of the "quantal" vowels [i], [u] and [a] whose properties will discussed below. In contrast, computer modeling shows that the SVTs of living primates, whose tongue are almost entirely within their mouths, inherently cannot produce quantal vowels because they cannot produce the necessary abrupt, midpointarae function discontinuities(Lieberman, Klatt and Wilson, 1969; Lieberman, Crelin and Klatt, 1972).

Figure 1 - Adult human supralaryngeal vocal tract (SVT) -- note that the tongue has an almost circular posterior (rear) contour. The “horizontal,”(SVTh)oral portion, and “vertical,” (SVTv) pharyngeal portion, have almost equal lengths. There is a natural discontinuity formed by the intersection of SVTh and SVTv that enables speakers to form abrupt changes in the cross sectional area of the human SVT at its midpoint.

Acoustic analyses of the vocalizations of non-human primates (e. g. Lieberman, 1968, Fitch, 1997, 2000; Rendall et al, in press) are consistent with modeling studies. Monkeys and apes produce schwa-like vowels (the vowel of the word "bub") because their tongue are positioned almost entirely in the mouth, characterize their vocalizations. One monkey species can produce two-formant frequency patterns that approximates a human [a] (Riede, et al., 2005). However, these vocalizations lack the third formant that would result from an [a]-like SVT; they appear to be generated by the laryngeal air sacs acting as resonators – a derived Diana monkey anatomical feature that has little relevance to the capabilities of ape and human vocal tracts that lack laryngeal air sacs (Lieberman,2006).[1]

What Makes Quantal Vowels Better than Other Vowels

Speech communication would be possible without quantal vowels. Indeed, as noted below, there would have been no selective advantage for retaining whatever mutations led to the evolution of the human SVT, unless some form of speech were already part of hominid culture. The term "quantal" was coined by Stevens (1972) to characterize speech sounds that have two useful properties. Quantal sounds have perceptually salient acoustic properties that can be produced with a certain degree of articulatory sloppiness

The task of speech production is simplified when it is possible to produce a stable acoustic signal without having to execute exceedingly precise articulatory maneuvers. The task of speech perception also is more robust if the resulting acoustic signals are maximally distinct. These criteria are captured by Kenneth Steven's (1972) "quantal factor." The quantal factor can perhaps be illustrated by means of the following analogy. Suppose that a trendy restaurant is to open. The owner decides to employ waiters who will signal the diners' order by means of acoustic signals. Shall he employ waiters equipped with violins or sets of handbells? If he wants to minimize the chance of errors he will opt for handbells which each produce a distinct acoustic signal without having to use precise manual gestures.

Figure 2 - Midsagital views of an adult human SVT for the quantal vowels [i], [a] and [u] and the resulting formant frequency patterns. Note the peaks in the frequency spectrum that follow from the convergence of two formant frequencies. The 10 to one discontinuity at the midpoint of the SVT allows speakers to be both imprecise and still generate vowels that have spectral peaks.[i]

Stevens demonstrated that the quantal vowels [i], [u], and [a] have perceptually salient acoustic "correlates" that can be produced while minimizing the need for precise motor control. Perceptual salience results from the convergence of two formant frequencies, yielding spectral peaks (Fant, 1960). For [i] the second and third formants, F2 and F3, converge at a high frequency; for [a] F2 and F1 converging at the midpoint of the frequency spectrum; for [u] F1 and F2 converge at a low frequency. A visual analogy may perhaps illustrate their communicative value. Using "quantal" vowels would be similar to communicating using flags that have brilliant saturated colors. Other vowels, whose formants do not converge produce formant patterns analogous to flags differentiated by pastel colors. Stevens demonstrated that if an abrupt area function discontinuity occurs at the midpoint of the SVT, the tongue can move as much as 1 cm back or forth without changing the formant frequencies appreciably. The exact position of the speaker's tongue with respect to the midpoint constriction for [i] does not have to be precise. Radiographic studies that track tongue movements confirm Steven's theory (Beckman et al., 1995).

Carre, Lindblom and MacNeilage (1995) using a different procedure, reached similar conclusions. Their SVT computer model "grew" a “vertical” pharyngeal portion (SVTv) that was equal in length to its “horizontal” oral cavity (SVTh) when directed at producing the full range of human vowels, delimited by [i], [u] and [a]. Radiographic and MRI studies show that the tongue body has a circular midsagital posterior contour and is almost undeformed when we produce vowels. Producing an [i] involves moving the tongue upwards and forward. An [a] can be produced by simply moving the tongue back and down (Russell, 1928; Chiba and

Kajiyama, 1941; Ladefoged et al., 1972; Nearey, 1979; Baer et al., 1991; Story, Titze, and Hoffman, 1996). The human tongue and those of virtually all mammals are "hydrostats" (Stone and Lundberg, 1996). Although muscular, the tongue can not be squeezed into a smaller volume as we produce different vowels. The intrinsic muscles of the tongue are sometimes bunched up when speakers produce an [u]or an [i] (Fujimura and Kakita, 1979. However, the shape of the tongue is usually a segment of a circular arc when vowels are produced.

Vocal Tract Normalization

The vowel [i] also facilitates estimating the length of a speaker's SVT. Longer SVTs yield lower formant frequencies than shorter SVTs for the same speech sound. Therefore, the absolute values of the formant frequencies of the same sound produced by different persons vary (Peterson and Barney, 1952; Hillenbrand et al., 1995). A perceptual “normalizing” process that takes account of SVT length is a critical step in speech perception.

The role of vocal tract normalization in speech became evident in the Peterson and Barney (1952) study of vowel formant frequencies and vowel perception. Figure 3 shows the Peterson and Barney plot of the vowel formants of seventy-six adult male, adult female and adolescent male and female speakers. The vowel symbols are plotted with respect to the values of their first and second formant frequencies. Each phonetic symbol plots F1 and F2 of the speakers’ vowels.

Figure 3

Plot of first and second formant frequencies derived by Peterson and Barney (1952) for the vowels produced by seventy-six different

speakers. Psychoacoustic studies show that F1 and F2 are sufficient to specify the vowels of English. The frequency of F2 is plotted with respect to the ordinate for each vowel token; the frequency of F1, with respect to the abscissa. The loops enclose 90 percent of the tokens produced by the speakers for each of the vowel categories of English.

The vowels' formant frequencies were measured from spectrograms of each speaker reading of a list of English words The spoken words were identified by listeners who had to identify each token without previously listening to a long segment of speech produced by each particular speaker. This was achieved by presenting a set of all of the words produced by ten different speakers in random order to the listeners. The listeners did not know whose voice or what word was coming next. A vowel symbol that falls into a loop marked with the same phonetic symbol signifies a token that was heard as the intended vowel. The loops on the plot in Figure 3 enclose the vowel tokens that made up 90 percent of the vowels that the speakers' intended to convey. The loops overlap even though they do not include 10 percent of the stimuli that fell into a nearby vowel class. The data, for example, show that many speakers [e] vowels had the same formant frequencies of other speakers' [I]s (the vowels of the words “bet” and “bit”) The general findings of the Peterson and Barney study were replicated, using computer-implemented formant analysis by Hillenbrand et al., (1995).

Human listeners can use different means to estimate a SVT’s length. Ladefoged and Broadbent (1957, for example, showed that the same tape-recorded word was perceived as “bit,” “bat,” or “but” depending on the average formant frequencies of a preceding

phrase. But we generally do not need to hear a person talking before we identify a word; there are immediate normalization cues in the speech signal. There were only two errors in 10,000 trials in identifying [i]s. in the Peterson and Barney study; [u]'s had 6 identification errors, whereas other vowels had high error rates. For example, [e] and [I] were confused hundreds of times. Nearey reasoned that the formant structure of [i] might make it possible for listeners to use it as an anchor point for vocal tract normalization.

Nearey predicted that a token of a formant frequency pattern in the [i] range will always be heard as an [i] produced by a SVT that had a particular length. If this were so, a listener would immediately "know" the length of a speaker’s SVT and would correctly associate formant frequency patterns with the vowel that the speaker intended to convey. In a controlled experiment, listeners first heard a "calibrating" [i] followed by a synthesized formant frequency pattern that could correspond to any vowel produced by either a short or a long SVT, followed by the same calibrating [i](Nearey, 1979, pp. 98-149). Nearey used two different calibrating [i]s, one that could be produced by an adult male's long SVT, one an [i] produced by an adolescent child's shorter SVT. Juxtaposed with the calibrating [i]'s were vowels having formant patterns that ranged over almost the total possible range of vowels for adult speakers and adolescent children. Listeners heard isolated sequences that had the for [i]-V- [i], where the [i]s were either long or short SVT [i]s and V the test stimuli. The listeners were told to identify each intermediate vowel “V.” The listeners were also asked to rate the naturalness of the vowel V of each [i-V-i] sequence that they heard. There were four categories of "naturalness judgment," from "OK" to "very bad."

The listeners' responses showed that they were "normalizing" SVTs using the single token of an [i], changing their identification of the identical formant frequency pattern when they heard it between long or short SVT [i]s. The listeners' "naturalness" responses demonstrated that they interpreted these synthesized speech stimuli using a mental procedure that "knew" the range of formant frequencies that can be produced by the calibrating [i]'s SVT length. For example, formant frequency patterns that could be produced by a short SVT were judged to be "natural" when they were embedded with an [i] produced by a short SVT but were judged to be non-speech stimuli when they were embedded with an [i] from a long SVt that inherently could not produce such high formant frequencies. The V vowels clearly were perceived in a "speech-mode," using neural processing that took account of the speech producing capabilities of the human SVT.

Other speech sounds can be used for vocal tract normalization (c.f. Lieberman (2006) for relevant studies), but the vowel [i] is an optimal calibrating sound. It's usefulness for SVT length estimation follows from its unique formant pattern (high frequency converging F2 and F3) and constraints on the vocal tract maneuvers that can be used to produce an [i]. Whereas alternate gestures can be used to generate the formant frequency patterns of virtually all other vowels, the tongue position and lip openings that generate an [i] are constrained (Stevens and House, 1955; Nearey, 1978). Speakers can protrude and constrict their lips to create the effect of having a longer SVT for most other vowels. Different tongue positions can be used for these vowels; speaker FSC in Nearey's (1978) study, for example, kept his tongue in almost the same position for almost all of his high F2 vowels, except for [i]. Alternate lip and larynx gestures generated his vowels’ formant

patterns. Fewer possibilities can generate the formant frequency patterns for an [i]. The tongue must be placed forwards and upwards to the point where turbulent noise is sometimes generated in the constricted oral passage necessary to produce an [i](Fant, 1960). The vowel [i] is an "honest" signal that specifies the speaker's actual SVT length. And it is one of the speech sounds that a non-human SVT cannot produce.

The neural mechanisms for perceiving formant frequencies and deriving SVT length appear to have a long evolutionary history Other species appear to use formant frequencies to estimate the size of a conspecific. Fitch (1997) used a simple metric obtained by the subtracting the frequency of F1 from F3 to estimate a monkey’s SVT length, which is highly correlated with its body weight and length. This metric works for other species (Fitch, 2000a) However, Fitch’s metric works only because these animal vocalizations are similar to the neutral “schwa” vowel of English in which F3 is approximately equal to 5(F1)(see the compilation for nonhuman primate species in Riede et al (2005)). If the same metric were applied to human speech it would yield different estimated SVT lengths for the same speaker, depending on the vowel analyzed since the formant patterns produced by humans diverge from the schwa vowel.

Why is Speech the Default Mode for Human Language?

Why do we talk? Why don’t we use manual gestures? The answer rests in some obvious factors, and one that became apparent through research conducted in the 1960s. Vocal communication frees a speaker’s hands, can occur in darkness, and doesn't require looking at the individuals who are signaling. A less obvious reason is the speed at which information can be transmitted by speech. Speech allows humans to transmit phonetic distinctions at

rates of up to 20 to 30 "segments" per second. Other auditory signals merge into a continuous buzz at rates exceeding 15 items per second. Speech achieves this rapid transmission rate because it is an "encoded" signal in which information is transmitted at the slower syllable rate then “decoded” into phonetic segments (Liberman et al., 1967).

For example, the formant frequency patterns that convey the “phonemes” of the word "cat" (approximated by the letters of the alphabet) are melded together into one syllable. As the tongue moves from the syllable-initial consonant, a formant frequency pattern is produced that transitions into that of the vowel, and then to the final consonant. Human speakers plan ahead. As you begin to say the word "too," your lips "round" (protrude and narrow) anticipating the rounded [u] vowel. Your lips are not rounded at the start of the word “tea”, because the following vowel is not rounded. The encoding differs somewhat from language to language (Lubker and Gay, 1982),and is acquired without conscious effort by children.

Choking and the antiquity of speech.

Speech must have been present in hominid species who lacked SVTs capable of producing quantal vowels because the human SVT increases the risk of choking to death on food lodged in the larynx. Palmer and his colleagues, reviewing studies of swallowing note that in contrast to nonhuman mammals:

> normal humans are at risk for inadvertently inhaling food particles both before and after swallowing. Indeed, obstruction of the airway by inhaled food is a significant cause of morbidity and mortality in otherwise healthy individuals." (Palmer et al., 1992).

Death resulting from a blocked larynx often is attributed to other causes, but tens of thousands of incidents of fatal choking have occurred (Feinberg and Ekberg, 1990). About 500,000 Americans suffer from swallowing disorders (dysphagia), and deaths from choking are the forth largest cause of accidental deaths in the United States (http;//www.nsc.org/library/report_injury_usa.htm). There would have been no reason for retaining the mutations that resulted in the human SVT, unless speech already was in place in hominids ancestral to humans before the evolution of the human SVT.

TRACING THE EVOLUTION OF THE HUMAN SVT

This brings us to the vexatious problem of reconstructing the soft tissue of the SVT of a fossil when all that remains are bones. Much attention has been given to the position of the larynx, which as we will see, can rule out hypothetical SVTs. However, studies of the ontogenetic development of the human SVT, discussed below, reveal other factors.

1- The skeletal structure that supports the roof of the mouth rotates towards the back of the skull, effectively shortening the mouth and the “horizontal component of the SVT, SVTh during the first two years of life; the human face is “flat” compared to prognathious present day apes and early hominids such as the Australopithecines (D. Lieberman, Ross and Ravosa, 2000).

2- The human tongue gradually descends down into the pharynx, changing its shape from a relatively long flat shape positioned almost entirely in the mouth to a massive form having a posterior rounded shape. This yields the 1:1 SVTh/SVTv proportions seen in Figure 1. This unique human developmental process is not complete until age 6-8 years (D. Lieberman and McCarthy, 1999). As the human tongue descends it carries the larynx down with it.

3- The human neck gradually lengthens (Mahajan and Bharucha, 1994). Neck length is critical since a larynx positioned below the

neck at the level of the sternum (collarbone) would make it impossible to swallow (Palmer et al., 2000; D. Lieberman et al., 2001).

Neanderthal speech.

As is the case in non-human primates throughout life, the tongue is positioned almost entirely in the mouth in human neonates. In the course of human ontogenetic development, the tongue moves down into the pharynx, carrying the larynx down with it. This process was first described by Victor Negus (1949)who thought that it reflected the:

> … recession of the jaws; there is no prognathous snout…The [human] tongue however retains the size it had in Apes and more primitive types of Man, and in consequence it is curved, occupying a position partly in the mouth and partly in the pharynx. As the larynx is closely approximated to its hinder end, there is of necessity descent in the neck; briefly stated the tongue has pushed the larynx to a low position, opposite the fourth, fifth and sixth cervical vertebrae. (Negus 1949, pp. 25-26)

Negus's inferences were correct insofar as extensive facial retraction occurs only in humans. As noted above, it has become clear that the process entails more than the recession of the jaws which occurs in the first two years of life. But these findings were almost 30 years in the future when, in 1971, Edmund Crelin and I attempted to reconstruct the SVT of the Neanderthal fossil found in the village of La Chapelle-aux-Saints (Boule, 1911-1913). We compared the skeletal features of the skull and mandible that support the soft tissues of the SVT in human newborns and the Neanderthal fossil. We noted the similarities that exist between the base of the skull and mandible of human newborn infants and the Neanderthal. A number of skeletal features were noted besides

basicranial flexure, which became the focus of many subsequent studies. These included skeletal features that support the muscles that move the tongue, such as the pterygoid process of the sphenoid bone, the total length of the basicranium and the distance between the end of the palate and the foramen magnum (into which the spinal column inserts). These basicranial lengths were similar in newborns and the fossil. On this basis the Neanderthal SVT was thought to be similar to that of a human newborn. Therefore, a range of SVT area functions similar to those of newborns in the cineradiographic study of (Truby, Bosma and Lind, 1965) was modeled using Henke's (1966) computer-implemented algorithm which established the relationships between SVT shapes and formant frequencies. The computed formant frequency vowels patterns were compared with those measured by Peterson and Barney (1952. Speech was possible since most vowel and consonant formant frequency patterns could be produced, but the formant frequency patterns that convey the "quantal" vowels of human speech could not be produced, owing to the reconstructed Neanderthal's tongue resting for the most part in the oral cavity. This precluded its producing the abrupt 10:1 area function SVT midpoint discontinuities necessary to produce quantal vowels.

Cranial base flexture.

A number of studies subsequent to the Lieberman and Crelin (1971) paper attempted to determine the probable SVTs of fossil hominids by establishing correlations between the cranial base angle and the SVT in living nonhuman primates and then making inferences based on this angle in a fossil. A fossil that had a shallow cranial base similar to that seen in living apes and human newborns presumably had a similar SVT, a fossil having a flexed adult human basicranial angle would have a human SVT. Similarities

between the embryonic and early stages of development have been used since Darwin (1859) to make inferences concerning evolution. Therefore, George (1978) studied the Denver series of cephalometric X-rays which tracked the development of basicranial skeletal features and the soft tissue of the SVT in children from age 3 months to adulthood (Maresh 1948; McCammon 1952). George correlated basicranial flexure with the occurrence of vowels that to her ears sounded like "quantal" vowels, such as [i]. An acute adult-like cranial base angle occurs at age two-years, when children appeared to produce quantal vowels. Since Stevens (1972) had shown that a SVT having adult proportions is necessary to produce these sounds, the conclusion was that the cranial base angle was an index of SVT proportions.

However, subsequent acoustic analyses showed that two-year old children do not produce the formant frequency patterns that specify quantal vowels. Buhr (1980) measured children’s vowel formant frequencies in the first years of life; they do not conform to those of adult speech. For example, the formant frequencies of a 64 week-old infant's vowels "heard" as [i] were actually those of [I] (the vowel of "bit"). But the difference in vowel quality is not apparent when listening to these utterances, even to trained phoneticians (e. g. Irwin, 1948). Patricia Kuhl and her colleagues in 1992 "solved" the mystery. When we listen to speech, a "perceptual magnet" pulls an ill-formed formant frequency pattern towards the ideal exemplar for the language that a person is exposed to in the early months of life. In effect, our speech perception system cleans up sloppy signals. The absence of computer-implemented digital image analysis technology in the 1970's precluded accurate measurements of tongue position by George; the perceptual magnet phenomenon documented by Kuhl and her colleagues was not apparent until almost two decades later. In

short, cranial base flexure, in itself, cannot be used to predict whether or not a fossil had an adult human SVT.

At the time, the supposed close relationship between SVT development and cranial base angle was shared by our and other research groups. Studies followed that linked the cranial base angle and the length of the basicranium (that indicates oral cavity length) with the SVTs of living non-human primates and fossil hominids (Laitman, Heimbuch and Crelin,1978, 1979; Laitman and Heim) Their conclusion was that Neanderthals and earlier fossil hominids earlier did not have human SVTs. The studies of Boe and his colleagues (Boe et al., 1999, 2001), which will be discussed below, reached an opposite conclusion.

Reconstructions of the SVTs of fossils based on cranial base angles are problematic. When Daniel Lieberman and McCarthy (1999) reexamined the Denver series they found that the tongue and larynx continue to descend after cranial flexure stabilized. The proportions of SVTh (the oral, horizontal segment) and SVTv (the pharyngeal, vertical segment), do not achieve their adult 1:1 proportion until age five to six years. Fitch and Giedd (1999) using MRIs, reached the same conclusion.

It’s the tongue, not the larynx.

The low position of the human larynx is a reflex of the human tongue reshaping and moving down into the pharynx. The position of the human larynx is closely coupled to tongue displacement (Negus, 1949; Bosma, 1975; D. Lieberman and McCarthy, 1999; Nishimura, et al., 2003). As the tongue descents down into the pharynx, it carries the larynx down with it. The descent of the tongue into the pharynx, its posterior circular shape and the right angle bend at its midpoint enables the human SVT to produce the major

midpoint area function discontinuities necessary for quantal vowels. Thus, despite the focus on the larynx in many studies on the evolution of speech, the descent and change in the tongue's shape is the key factors in both the development and evolution of the human SVT (Lieberman (1984, pp. 276-280).

Studies of species whose tongues are positioned in their mouths show that their vocalizations are limited to the schwa vowel. Fitch's (1997, 2000a) data, for example, shows that this is the case. Non-human SVT phonetic limitations characterize the deer vocalizations studied by Fitch and Reby (2001). Although the deer have low larynges, their tongues remain anchored in their long mouths. That is also the case for lions whose larynges transiently descend as they roar; an elastic membrane links the larynx to a tongue anchored in the mouth (Weisengruber.et al., 2002). The larynges of young chimpanzees descend somewhat through elongation of the distance between the hyoid bone and the larynx, but their tongues do not descend (Nishimura et al.,2003). In short, in itself, a low larynx is not an indicator of potential phonetic ability. Claims such as Fitch (2000b) that the human SVT evolved to produce lower formant frequencies by laryngeal descent (providing a false vocal impression of a larger body), cannot account for the evolution of the species-specific human SVT which involves the descent of the tongue into the pharynx.

Recent Incorrect Inferences Concerning Neanderthal SVTs

The biological mechanisms that regulate the descent of the tongue and reshaping of the human are presently unknown and tongue position and shape cannot be inferred from the basicranial angle. Boe and his colleagues in 1999 and 2002 nonetheless base their Neanderthal reconstruction on the cranial base angle of the La Chapelle-aux-Saints fossil as reconstructed by Heim (1989). The

basicranial flexure of Heim's Neanderthal skull reconstruction is within the human range, but that does not signify an adult human SVT. Although the D. Lieberman and McCarthy (1999) and Fitch and Giedd (1999) studies are cited by Boe and his colleagues, they ignore their findings and fit a SVT having the adult human proportions noted by Honda and Tiede (1998) to the fossil.

The relationships that hold between skulls, jaws and soft tissue noted by Honda and Tiede (1998) hold for adult humans; they do not apply to young children, human neonates, apes or monkeys. Genetic evidence (Krings et al, 1997; Ovchinnikov et al., 2000) show that Neanderthals diverged from humans about 500,000 years ago. Their skeletal morphology differs from that of modern humans (Howells, 1976, 1989; D. Lieberman, 1995). In short, adult Neanderthals are not genetically or morphologically similar to modern human adults. Adult human SVT morphology thus cannot arbitrarily be bestowed to Neanderthals. Nonetheless, Boe and his colleagues model the SVT shapes that adult human speakers use to produce vowels. Not surprisingly, these human vocal tract configurations produce the full range of human vowels. Boe and his colleagues also model a putative human infant SVT in their 1999 and 2002 papers, that does not resemble any newborn SVT documented by Negus (1949), Truby, Bosma and Lind, 1965; Bosma (1975) or anyone else. Its SVTv/SVTh ratio is close to that of the five to six year-old children documented in the Lieberman and McCarthy (1999) and Fitch and Giedd (1999) studies. Similar flaws mark other studies that proposed human SVTs for Neanderthals, discussed in Lieberman (1984, 2000, 2006).

When did a fully modern human SVT evolve?

A SVT that can produce the full range of human speech must have 1:1 SVTh to SVTv proportions. If SVTh is long, as is the case

for Neanderthals, than SVTv must also be long. But the anatomy involved in speech (tongue, hyoid bone, and larynx) has a more “primitive,” basic function -- eating. The hyoid which supports the larynx, moves upwards and forwards about 13 mm, opening the espohagous and placing the larynx into a position in which food will not fall into it while swallowing (Ishida, Palmer, and Hiiemae, 2002). A larynx in the neck can execute these maneuvers. However, if the cricoid cartilage of the larynx were placed in the chest, the sternum bone would make it impossible to execute these movements. The swallowing "pattern" generator” - the movements that are involved in swallowing are similar in humans and apes (Palmer et al., 2002). No human or ape descended from our common ancestor has a larynx in its chest because they would not be able to eat.

We can determine whether Neanderthals and other fossil hominds could have had 1:1 SVTh to SVTv proportions by examining their basicrania, which provides a measure of SVTh and their cervical vertebrae, which provides a measure of the length of their necks. McCarthy,et al, forthcoming)determined these metrics for a sample of 62 specimens of *Pan troglodytes,* the WT 15000 fossil *Homo ergaster,* three Neanderthal fossils*,* 82 specimens of *Homo sapiens*,including the Middle Paleolithic Skhul V fossil, eight Upper Paleolithic fossils, and 73 contemporary humans from seven different populations. The data show that Neanderthal neck lengths were too short to have fully human SVTs. McCarthy and his colleagues arrive at a Neanderthal neck length estimate of 120 mm in contrast to the 134 to 127 mm averages for two modern human samples; the short neck and long Neandertal SVTh would place the cricoid cartilage behind the sternum. Fully human speech would be possible, but the hypothetical Neanderthal would be unable to eat. (A similar conclusion was reached in Lieberman (1984, pp. 290-296).

Surprisingly, a similar constraint rules out a fully human SVT in the middle Pleistocene fossil Skhul V (McCowan and Keith, 1939), which has often been thought to be fully modern. McCarthy and his colleagues estimate the cervical spine length of Skhul V to be 109 mm, at the bottom of the adult modern human range. Skhul V's SVTh is relatively long. Therefore, its short neck precludes its having a fully human vocal tract with 1:1 SVTh to SVTv proportions. Fully modern speech anatomy is not evident in the fossil record until the Upper Paleolithic, about 50,000 years ago

THE NEURAL BASES OF SPEECH

It is clear that human speech entails having neural capabilities that are absent in closely related living species. Although a chimpanzee's SVT would suffice to establish vocal language, they cannot talk. This despite the fact that acoustic analyses (e.g., Lieberman, 1968) reveal "bound" formant frequency patterns in chimpanzee calls similar to those that convey different words in human speech. These sounds could be used to differentiate words if the chimpanzees could voluntarily reorder the motor commands used to generate them. Chimpanzees could establish "protospeech," producing everything save quantal sounds if they were able to freely reiterate – to voluntarily reorder and recombine the motor commands underlying speech. Chimpanzees calls in the state of nature appear to be stereotyped and fixed (Goodall, 1986). The neural circuits that confer the reiterative abilities necessary for human speech appear to be absent in chimpanzees and other non-human primates.

The reiterative quality of these human neural circuits extends to other aspects of behavior, including syntax. I shall briefly review studies that support this claim; these studies also

show that traditional Broca-Wernicke "language organ" theory is wrong. Cortical-striatal-cortical neural circuits that include the basal ganglia appear to regulate motor control, syntax and cognition. The subcortical basal ganglia constitute a "sequencing engine" that can reiterate motor commands stored as “motor pattern generators” in other parts of the brain. The basal ganglia through different anatomically segregated neural circuits also reiterate cognitive "pattern generators," conferring cognitive flexibility and take part in associative learning. The evolutionary significance of the regulatory FOXP2 gene, which has erroneously been identified as a "language gene," rests in the fact that it governs the embryonic development of the basal ganglia and other subcortical elements of these neural circuits. Fuller accounts of these issues are presented in Lieberman (2000), (2002) (2006) and the studies noted below.

Neural Circuits

Complex brains contain many distinct neuroanatomical structures that in normal circumstances process particular tactile, visual, or auditory stimuli, while other structures and cortical regions perform local operations that regulate aspects of motor control or hold information in short-term (working) memory, etc. (e.g. Marsden and Obeso, 1994; Mirenowicz and Schultz, 1996; Monchi et al., 2001; Polit and Bizzi, 1978; Sanes et al., 1995). However, an isolated structure or cortical area usually does not by itself regulate a complex behavior. Individual neural structures generally contain many anatomically segregated groups, “populations," of neurons that carry out a particular “local” operation. The local processes do not constitute an observable behavior. The neuronal population that carries out a local process is linked to, "projects" to, anatomically distinct neuronal populations in other regions of the brain. The series of linked

neuronal populations form a neural "circuit." The circuit constitute the brain basis of an observable aspect of behavior -- walking, talking, striking the keys of a computer’s keyboard and so on. Moreover, within a given neural structure, distinct anatomically segregated neuronal populations may occur that project to neurons in different brain structures, forming multiple circuits that each regulate some other behavior.

As Dobzhansky (1973) put it, “Nothing in biology makes sense except in the light of evolution.” Neural structures that were initially adapted to control one function took on “new” tasks. Seen in this light, the local motor sequencing operations in the subcortical basal ganglia discussed below appear to be precursors for similar operations in cognitive domains. As we shall see, the basal ganglia can alter a motor act when circumstances dictate by switching from one “motor pattern generator” to another more appropriate one. During a thought process they can switch from one “cognitive pattern generator” to another (Graybiel, 1997).

For example, within the putamen, a subcortical basal ganglia structure, anatomically segregated populations of neurons exist that form part of a system that sequences the motor sub-movements that together constitute an overt movement of a monkey's hand, a rat's grooming sequence, or a person's walking or speaking (Aldridge et al., 1993; Cunnington et al., 1995; Lieberman, 2000; Marsden and Obeso, 1994). The putamen, in itself, is not the "seat" of these motor acts; it acts as a device that, in essence, connects the sub-movement pattern generators to areas of motor cortex. Distinct, anatomically segregated neuronal populations in the putamen project through other subcortical structures to cortical areas implicated in higher cognition, comprehending the meaning of a sentence, attention, and reward-based learning (e.g.

Alexander, Delong and Strick, 1986; Alexander and Crutcher, 1990; Cummings, 1993; Graybiel, 1995, 1997; Kimura, Aosaki and Graybiel, 1993; Marsden and Obeso, 1994; Middleton and Strick, 1994).

Experiments-in-Nature and the Traditional Broca-Wernicke Model.

The study of the neural bases of human language began with "experiments-in-nature" that produced "aphasia", permanent loss of linguistic abilities, after parts of the brain were destroyed by accidents, strokes, or other pathologies. Experiments-in-nature still are germane to the brain-language question, particularly when their findings are integrated with tracer, imaging and electrophysiological studies. Paul Broca's (1861) observations arguably rank with the most influential "experiments in nature." However, the interpretation of brain-behavior relationships presented here is quite different from Broca's.

Broca’s patient, "Tan," had a series of strokes. The strokes had caused extensive brain damage including, but not limited to one part of the brain, "the third frontal convolution" an anterior (front) area of the cortex. Tan's had limited speech ability and only uttered the syllable "tan". Broca perhaps influenced by earlier phrenological theories (Spurzheim, 1815), concluded that damage to this cortical region, which includes Broca's area, was the basis of the patient's speech deficit. If one’s model of the brain is that discrete localized regions regulate observable complex behavior, it follows that destroying a region should disrupt a particular aspect of behavior Overlooked was the fact that Tan also had extensive subcortical damage and nonlinguistic motor impairments. Wernicke in 1874 found that patients who had suffered damage in the posterior left hemisphere had difficulty comprehending speech. Again, Wernicke's localized receptive linguistic ability to this neocortical area. Since language

involves both comprehending and producing speech or alternate phonetic systems such as writing or sign language, Lichtheim (1885) proposed a cortical pathway linking Broca's and Wernicke's areas. According to this model, spoken language is perceived in Wernicke's area, a posterior temporal region associated with auditory perception. A cortical pathway then transmits information to Broca's region, which is adjacent to cortical areas implicated in motor control.

Although the Broca-Wernicke model has the virtue of simplicity, it is at best incomplete. The behavioral deficits of Broca's aphasia are not limited to speaking; the linguistic deficits involve difficulty comprehending distinctions in meaning conveyed by syntax and word-finding difficulties (Blumstein, 1995). Patients also suffer from cognitive deficits. Kurt Goldstein (1948) characterized Broca's aphasia as "loss of the abstract capacity," and noted an inability to adapt to changing circumstances. Contemporary clinical evidence shows that permanent loss of language does not occur absent subcortical damage, even when Broca's or Wernicke's areas have been destroyed. For example, although MRIs showed almost complete destruction of Wernicke’s area in a 60 year-old patient, he made a full recovery; no subcortical damage was apparent (Lieberman, 2000, pp. 101-102) Moreover, damage to subcortical structures, sparing cortex, can produce aphasic syndromes. Doubts had been expressed in the early years of the twentieth century, but computer aided tomography (CT) scans and magnetic resonance imaging (MRI) now provide information on the nature and extent of brain damage that produces permanent language loss. Aphasia does not occur unless subcortical damage is present (Stuss and Benson, 1986; Dronkers et al., 1992; D'Esposito and Alexander, 1995). Patients having extensive damage to Broca's area generally recover unless subcortical damage also occurs.

Other studies show that subcortical damage that leaves Broca's area intact can result in Broca-like speech production and language deficits (e. g. Naeser et. al., 1982; Benson and Geschwind, 1985; Alexander, Naeser and Palumbo, 1987).

Alexander and his colleagues (1987), for example, noted the subcortical locus of aphasias, reviewing l9 cases of aphasia that resulted solely from subcortical lesions. The language deficits ranged from fairly mild impairment in a patient's ability to recall words, to "global aphasia" in which a patient produced very limited speech. In general, the severest language deficits occurred in patients who had suffered the most extensive subcortical brain damage and damage to the internal capsule (the nerve fibers that project to the cortex. Subsequent studies rule out damage to the internal capsule as causing aphasia. Deliberate surgical lesions of the internal capsule aimed at mitigating obsessive-compulsive behavior do not induce aphasia (Greenberg, Murphy and Rasmussen, 2000).Damage to the basal ganglia from strokes in the medial cerebral artery which passes through them may be the locus of Broca’s aphasia. As D'Esposito and Alexander (1995) in their study of aphasia conclude, it is apparent,

> That a purely cortical lesion--even a macroscopic one--can produce Broca's or Wernicke's never been demonstrated. (1995, p. 41)

Cortical-striatal-cortical circuits.

The basal ganglia are subcortical structures located deep within the brain. They can be traced back to anurans similar to present day frogs (Marin, Smeets and Gonzalez, 1998). The striatal component of the basal ganglia includes the caudate nucleus and the lentiform nucleus. The lentiform nucleus itself consists of the putamen and globus pallidus. The putamen receives sensory

inputs from most parts of the brain. The globus pallidus is an output structure receiving inputs from the putamen and caudate nucleus. The caudate nucleus, putamen, and globus pallidus are interconnected and form a system with close connections to the substantia nigra, thalamus, other subcortical structures and cortex. The thalamus, in turn,connect to different cortical areas. The connections with cortex are complex (Alexander, Delong and Strick, 1986; Parent, 1986; Alexander and Crutcher, 1990; DeLong, 1993; Marsden and Obeso, 1994; Middleton and Strick, 1994.

Disruptions in behavior seemingly unrelated such as obsessive-compulsive disorder (Greenberg, Murphy and Rasmussen, 2000), schizophrenia (Graybiel, 1997) and Parkinson's Disease (Jellinger, 1990) derive from the disruption of neural circuits linking cortical areas with the basal ganglia. Behavioral changes usually attributed to frontal lobe cortical dysfunction can be observed in patients having damage to basal ganglia (e. g., Cummings and Benson, 1984; Flowers and Robertson, 1985; Alexander, Delong and Strick, 1986; Lange et al., 1992; Delong, 1993;).

Cummings in his 1993 review article identifies five parallel basal ganglia circuits which are involved in motor control, cognition, attention and other aspects of behavior. The circuit (probably circuits) projecting to the dorsolateral region of prefrontal cortex (a frontal region of the cortex) is associated with cognitive behavior. Tracer studies confirm these circuits. Traditional tracer studies entail injecting substances into living animals that attach themselves to the outputs of neurons projecting to other neurons forming neural circuits. Post-mortem sectioning, staining, and microscopic examination then reveal the neural pathways. Tracer studies of monkey brains confirm that the striatal basal ganglia (the caudate nucleus and putamen) support circuits that project to cortical areas associated with motor

control and cognition (Alexander, Delong and Strick, 1986; Middleton and Strick, 1994; Graybiel et al., 1994; Graybiel, 1995, 1997). Noninvasive Diffusion Tensor Imaging (DTI) techniques that are based on MRI technology, show similar neural circuits in humans (Lehericy et al. 2004).

Neurodegenerative Diseases

Parkinson's (PD) damages the basal ganglia, mostly sparing cortex (Jellinger, 1990). The primary deficits of PD are motoric; tremors, rigidity, and movement disruptions occur. In PD, speech production deficits occur similar in nature to those occurring in Broca's aphasia. Patients have difficulty sequencing the lip, tongue and laryngeal maneuvers necessary to differentiate "stop" consonants." Stop consonants are produced by momentarily obstructing the SVT with the lips (for [b] and [p]) or tongue (for [d], [t],[g] and [k]). The lips or tongue then open the SVT producing a momentary “burst,” an abrupt pulselike acoustic signal. The larynx must then produce phonation keyed to the burst. Phonation must occur within 20 msec. from the burst for the English "voiced" stops [b], [d], and [g] (the initial consonants of the words "bad," "dab," and "god"). Phonation must be delayed, usually for at least 60 msec. for the English "unvoiced" stops [p], [t], and [k] (the initial consonants of "pad," "tab," and "cod"). This phonetic distinction, which entails controlling the sequence of gestures between tongue or lips and the muscles of the larynx, was termed "voice-onset-time" (VOT) by Lisker and Abramson (1964).

Similar VOT distinctions differentiate the stop consonants of all human languages analyzed to date. (Many languages also differentiate words by means of “prevoiced” stops in which voicing starts before the burst.) Acoustic analyses show that a breakdown in regulating VOT is the most symptomatic speech deficit of

Broca's aphasia (Blumstein et al, 1980; Baum et al, 1990)and in PD (Lieberman et al.,1992, 2000). In contrast, formant frequency patterns that reflect SVT maneuvers are generally preserved in both Broca's aphasia and PD (Blumstein, 1994; Lieberman, 2000).

As is the case for Broca's aphasics (Blumstein, 1995), PD patients (Illes et al., 1988) can have difficulty producing sentences that have complex syntax. PD patients likewise have difficulty comprehending sentences that have moderately complex syntax as well as long sentences that tax the brain's computational resources (e.g., Lieberman, et. al., 1992; Natsopoulos et al., 1993; Grossman et. al., 1991, 1993; Lieberman, 2000; Hochstadt, 2004). As PD progresses, dementia occurs, different in kind from Alzheimer's (Cummings and Benson, 1984). Afflicted patients retain semantic and real-world knowledge but are unable to readily form or change cognitive sets (Flowers and Robertson, 1985; Cools et al., 2001). These seemingly unrelated deficits derive from the "local" operations performed by the basal ganglia in the cortical-striatal-circuits that regulate these aspects of behavior.

Basal Ganglia Operations:

The basal ganglia operations characterized by Graybiel (1995, 1997, 1998) involve both “motor pattern generators” and “cognitive pattern generators.” In the era before medication with Levadopa was used to treat Parkinson's Disease, thousands of operations were performed. The effects were reviewed in a seminal paper by Marsden and Obeso (1994). They note that the basal ganglia have two different motor control functions.

First, their normal routine activity may promote automatic execution of routine movement by facilitating the desired cortically driven movements and suppressing unwanted muscular

> activity. Secondly, they may be called into play to interrupt or alter such ongoing action in novel circumstances … they respond to unusual circumstances to reorder the cortical control of movement. (Marsden and Obeso, 1994, p. 889)

Marsden and Obbeso conclude that,

> Perhaps the basal ganglia are an elaborate machine, within the overall frontal lobe distributed system, that allow routine thought and action, but which responds to new circumstances to allow a change in direction of ideas and movement. Loss of basal ganglia contribution, such as in Parkinson's disease, thus would lead to inflexibility of mental and motor response..." (1994, p. 893).

Neuroimaging Studies

Brain imaging studies of human subjects confirm this inference. The event-related functional magnetic resonance imaging (fMRI) study of Monchi et al. (2001) shows the role of basal ganglia when a person shifts cognitive sets. Brain activity was monitored in neurologically intact subjects in a version of the Wisconsin Card Sorting Test (WCST), which evaluates a person's ability to form and shift cognitive criteria. Subjects had to sort cards by matching the images on them to the colors, shapes, or number of images on “match” cards. As predicted, neural circuits involving prefrontal cortex and basal ganglia were activated throughout the test. Bilateral activation was observed in prefrontal cortex, basal ganglia and thalamus. Dorsolateral prefrontal cortical areas were active at the points where the subjects had to relate the current match with earlier events stored in working memory. A cortical-striatal circuit involving a different cortical area, (the mid-ventrolateral prefrontal cortex), caudate nucleus, putamen and thalamus was active when subjects had to shift to a different matching criterion. Increased activity

occurred in the putamen during these cognitive shifts. The behavioral study of Scott and his colleagues (2002) complements these findings. A comprehensive set of cognitive tests that assess "frontal lobe" functions such as planning as well as tests of memory were administered to PD patients who had undergone neurosurgery that produced precise bilateral lesions of the internal output pathway, of the globus pallidus. The sole deficits occurred on the Wisconsin Card Sorting Test, where the subjects were unable to shift the matching criterion as the test progressed.

Stowe et al. (2004) used PET imaging of neurologically intact subjects in a sentence comprehension study that involved a form of set shifting. The basal ganglia to dorsolateral prefrontal cortex circuit was active when subjects have to change their interpretation of an ambiguous sentence, confirming that basal ganglia cognitive set shifting also manifests itself in language. Other neuroimaging studies show basal ganglia as well as cortical activity during sentence comprehension and word retrieval tasks (Klein et al., 1994; Kotz et al., 2003; Rissman, Eliassen and Blumstein, 2003).

The focus on subcortical structures here should in no way imply that cortex is irrelevant. The imaging studies noted above and many other studies show that Broca’s area is active when a person listens to speech, when a person recalls a word as well as when the meaning of a sentence or the when a listener identifies the emotional content of a sentence. Cortical areas in both hemispheres of the cortex are active in these tasks, including the right hemisphere homologues of Broca’s and Wernickes’s areas and prefrontal areas that are not traditionally associated with language.(Just et al.,1996) The absence of basal ganglia activity in other imaging studies may reflect "region of interest" (ROI)

procedures that did not look for subcortical activity during linguistic tasks.

Electrophysiologic studies that monitor brain activity in monkeys by means of exceeding fine “microelectrode” probes show that the basal ganglia perform similar functions (reviewed in Graybiel, 1995, 1997, 1998) as well as in other mammals. When the basal ganglia of rats are destroyed they are able to execute the individual submovements that when linked together would constitute a grooming sequence (Berridge and Whitshaw, 1992), but they cannot perform the complete grooming sequence. Electrophysiologic studies of the rodents' basal ganglia neurons show firing patterns that sequentially inhibit and release submovements to the motor cortex, thereby stringing them into a grooming sequence (Aldridge et al., 1993)

A laboratory Called Mount Everest

It generally is not possible to compare the behavior of human subjects before and after an insult to the brain. Nor is it ethically justifiable to test theories by placing subjects in a situation that might harm their brains. However, mountain climbers who are determined to reach the summit of Mount Everest, provide a unique, ethically sound situation in which the effects of basal ganglia dysfunction on motor control, language, cognition and other aspects of behavior can be determined. The cognitive abilities of individual subjects can be assessed before and after hypoxic insult to their brains, allowing the assessment of subtle as well as profound impairment. Everest provides an opportunity to focus on basal ganglia function because hypoxia (oxygen deficits) commonly occurs as mountain climbers ascend. Metabolically active neural structures such as the basal ganglia are particularly sensitive to hypoxia (Inoue et al., 1992; Burke et al., 1994).

Independent studies show that the globus pallidus is extremely sensitive to hypoxic damage (Laplane et al., 1984, 1989; Strub, 1989). MRI imaging confirm bilateral lesions localized to globus pallidus after exposure to altitude; the lesions produce subcortical dementia and aphasia (Jeong et al. 2002; Chie et al., 2004).

A series of experiments (Lieberman et al., 1994, 2005) shows that speech production deficits similar to PD as climbers become more hypoxic as they ascend to higher altitudes. VOT sequencing is impaired and their speech slows down as the duration of their vowels increases. Cognitive tests such as the WCST, administered at successively higher altitudes show that the set-shifting performance declines. Sentence comprehension also slows down and error rates increase. In extreme cases, hypoxic climbers exhibiting profound speech and set-shifting errors did not adapt their behavior to changing life-threatening events. Shifts in personality also occur, similar in nature to those reported by Cummings (1993) for damage to cortical-striatal-cortical circuits.

Motor Control and Reiterative Ability

Many linguists (e.g. Jakendoff, 1994; Chomsky, 1999) still hold to the view that human language is so unique that it bears little relation to the manner in which any other animal communicates or thinks. Chomsky, whose focus has been on syntax for many years, has consistently argued that human syntactic ability involves some unique feature whose scope is restricted to language and language alone. As noted earlier, the most recent candidate (Hauser, Chomsky and Fitch, 2002), is a "narrow faculty of language" (FNL) that confers recursion. Chomsky's initial (1957) generative syntactic theory proposed that the relative clause in the sentence *I saw the boy who was wearing a sweater,* was the end product of a process in which a hypothetical underlying sentence,

The boy was wearing a sweater, had been inserted into the frame of the carrier sentence “I saw the boy.” Subsequent hypothetical “transformational” rules of the generative grammar then rewrote the resulting string of words to yield the sentence, *I saw the boy who was wearing a sweater,* that would actually be heard or read. Traditional grammars would straightforwardly characterize the actual, observable sentence as containing a relative clause.

In Chomsky’s (1999) current “minimalist” grammar the syntactic rule “merge” recursively inserts sentences and other syntactic units into the framework of a carrier sentence; the minimalist syntactic rule “move’ then rewrites the resulting string of words to yield the sentence that one actually hears or reads. The reiterative function of the basal ganglia includes reordering and replicating cognitive pattern generators (Graybiel, 1997). The cognitive pattern generator that elicits the relative clause, *who was wearing a sweater,* would simply be inserted into the frame of the carrier sentence.

In short, the basal ganglia sequencing engine can form a potentially infinite number of different sentences by reordering, recombining and modifying a finite set of words using a finite set of linguistic "rules." Reiteration can account for the sentences that we actually hear, inserting a relative clause, a prepositional clause, whatever, into a carrier phrase. In principle, the linguistic process is no different than inserting the dance instruction “allemande right” into a square dance, or Mozart inserting yet another variation into a rondo. Reiteration can also account for the formal phonologic operations used by linguists to describe word-level phonologic processes such as the formation of “regular” English plural nouns by adding the sounds coded by the suffix “s,” e. g. book versus books.

Chomsky and his colleagues are correct in proposing processes that can generate a potentially infinite number of sentences or words from a finite set of words and rules. However, as we have seen the ability to reorder and recombine a finite set of elements to form an infinite set of actions is a key feature of speech motor control, other motor acts and aspects of “nonlinguistic” cognition such as changing the direction of one’s course of action, or changing the criterion by which one categorizes objects.

Many linguists may argue that language is quite different than motor control. Forming different grammatical sentences entails more than inserting a phrase or word or simply changing word order. The semantic-syntactic constraints on the words in any dictionary, including that in your brain, must be taken into account. Different verbs, for example, have particular constraints (the linguistic term generally used is "argument structure"). For example, the ungrammatical sentence "I wished Ann." violates a constraint because the verb "wish" cannot refer to an object, whereas "I kissed Ann." is acceptable. Motor control entails similar, indeed more complex constraints. As the basal ganglia release and inhibit successive pattern generators, these constraints come into play. Consider walking, which involves a sequence of sub-movements. Heel-strike, one component of walking can only be executed after the motor pattern generator that swings the lower leg forward. Nor can the pattern generator that locks your legs in place while you standing still be followed with heel strike. Running, which appears to have shaped human evolution, (Bramble and D. Lieberman, 2004) requires exceedingly rapid and precise control of a different set of pattern generators. If walking or running seem too simple, consider the set of sequential motor commands involved in baseball, playing the violin, or

dancing the tango. In short, motor pattern generators have "argument structures."

The FOXP2 Gene.

The FOXP2 gene undoubtedly is not the only regulatory gene involved in the evolution of human language. Moreover, it is not a "language" gene since it governs the embryonic development of neural structures that regulate motor control, other aspects of cognition, and emotional regulation, as well as the development of lung tissue and other structures. However, studies of the FOXP2 gene provide some insights on the evolutionary history of the human brain.

The discovery of FOXP2 results from a sustained study of a large extended family marked by a genetic anomaly. A "syndrome," a suite of speech and orofacial movement disorders, and cognitive and linguistic deficits occurs in afflicted members of the KE family (Vargha-Khadem et al., 1995, 1998; Lai et al., 2001 Watkins, et al., 2002). Afflicted individuals are not able to protrude their tongues while closing their lips; they have difficulty repeating two word sequences. On standardized intelligence tests, they have significantly lower scores than their non-afflicted siblings. Some afflicted individuals had higher non-verbal IQ scores than unaffected members of the KE family, which leads some investigators to conclude that FOXP2 does not affect intelligence. However, as the different non-verbal IQs for the non-affected members of the KE family show, intelligence derives from the interaction of many neural systems and life's experiences. It is impossible to know what the non-verbal IQs of an affected individual would have been, absent the genetic anomaly, but the low mean 86 non-verbal IQ of the affected members (with a range of 71-11), versus a mean of 104 (with a range of 84 to 119) for

unaffected family members, suggests FOXP2 anomalies being responsible for generally lower intelligence.

MRI imaging of affected family members shows that the caudate nucleus is abnormally small bilaterally, while the putamen, globus pallidus, angular gyrus, cingulate cortex and Broca's area are abnormal unilaterally. (Vargha-Khadem et al., 1995, 1998). Watkins et al.(2002) note that reduced caudate nucleus volume was " significantly correlated with family members' performance on a test of oral praxis, non-word repetition, and the coding subtest of the Wechsler Intelligence Scale." fMRI studies that compare afflicted members of the KE family with both their "normal" siblings and age-matched controls show that under-activation occurs in the putamen, Broca's area and its right homolog (Liegeois, et al. 2003), which is what would be expected in neural circuits connecting the striatum and Broca's area (Lehericy et al., 2004). The pattern between of neural anomalies and behavioral deficits is similar to those seen in individuals afflicted with PD, hypoxia and lesions in basal ganglia.

This constellation of neural anomalies and behavioral deficits results from a dominant point mutation mapped to chromosome 7q31 in the FOXP2 gene (Fisher et al., 1998; Lai et al. 2001). Lai and her colleagues determined the neural expression of FOXP2 during early brain development in humans, and the mouse version (foxp2), in mice (Lai et al. 2003) -- mammalian "end points" separated by 75 million years of evolution (Mouse genome sequencing consortium, 2002). The gene encodes a protein that regulates the expression of other genes during embryogenesis. Mutations to other similar genes have been implicated in a number of developmental disorders. In the case of family KE, the mutation

changes an amino acid, apparently leading to protein dysfunction. The similar areas of expression that indicate where the gene is active in both the human and mouse brain include structures in the cortical-striatal-cortical circuits that regulate motor control and cognition -- the thalamus, caudate nucleus and putamen as well as the inferior olives and cerebellum. These structures are all intricately interconnected. Independent evidence shows that foxp2 in other mammals is expressed in the putamen as well as the caudate (Takahashi et al., 2003).

The FOXP2 gene provides a means to date the evolution of the human brain and the emergence of fully human speech capabilities. Despite the high degree of similarity there are important distinctions between the mouse, chimpanzee and human versions. The mouse and human versions are separated by three mutations. The chimpanzee and human versions are separated by two mutations. Enard et al. (2002), using the techniques of molecular genetics, estimate that the human form appeared fairly recently, somewhere in the last 100,000 years, 100,000 years being in the time frame (Stringer, 1998) associated with the emergence of anatomically modern *Homo sapiens.*

Walking, Running and the Antiquity of Speech

One point concerning the evolution of human speech deserves more emphasis, its antiquity. The Lieberman and Crelin (1971) Neanderthal study is often cited to support claims that speech evolved abruptly at a recent date. Boe et al. (1999, 2002), claim that we concluded that Neanderthals were a "speechless species." However, this was not our conclusion, what we wrote was that Neanderthals represent:

> ...an intermediate stage in the evolution of language. This indicates that the evolution of language was gradual, that it

> was not an abrupt phenomenon. The reason that human linguistic ability appears to be so distinct and unique is that the intermediate stages in its evolution are represented by extinct species (Lieberman and Crelin, 1971, P. 221).

Some form of speech must have been in place in the archaic hominids ancestral to both humans and Neanderthals. There would have been no selective advantage for retaining of the mutations that yielded the species-specific human SVT at the cost of increased morbidity from choking, unless speech was already present. The question is when?

The basal ganglia dysfunction that is the proximate cause of PD impairs walking. PD patients have difficulty executing the internally guided sequential movements involved in walking. The Hoehn and Yahr (1967) diagnostic scale for PD is a measure of upright balance and locomotion. Running is impossible. As we have seen, the subcortical structures whose expression is regulated by FOXP2, the basal ganglia and cerebellum, play a critical role in motor control, motor learning as well as in cognition. Learning to execute a motor sequence involves activity in these subcortical structures as well as prefrontal cortex (e. g. Kimura, Aosaki and Graybiel, 1993; Thatch, 1996).

Selection for walking, starting from the base apparent in present day chimpanzees who can walk for limited periods, perhaps was the "start-point" for the evolution of human speech, language and cognition. The evolution of the genus Homo was marked by adaptations for endurance running (Bramble and D. Lieberman, 2004), which places still further demands on the basal ganglia sequencing engine. Lacking more data, we can only speculate that a neural substrate permitting voluntary speech motor control was in place in early Homo erectus. Further selection for speech production may

have resulted in the human form of FOXP2 and the motor, cognitive, and linguistic abilities of contemporary humans. Developmental-neurophysiologic studies comparing the development of walking and speech may move this proposal beyond speculation.

Putting Anatomy and the Brain Together

The findings discussed here, concerning the evolution of human speech anatomy and the human brain, point to the same conclusion. The evolution of speech was driven by Darwinian Natural Selection, the opportunistic use of existing structures adapted for another purpose, and mutations on regulatory genes that have far reaching consequences. Contemporary human speech and cognitive capabilities, including enhanced syntactic and lexical abilities, are species-specific properties of *Homo sapiens* which derive from anatomy and neural mechanisms that appear to have coevolved. The FOXP2 gene clearly is implicated in the formation of neural circuits that regulate human cognitive and motor capacities. Natural selection acting on the mutations that yielded its human form would have enabled rapid, encoded speech, in turn enhancing the selective value of the mutations that shaped the modern human vocal tract. These events which led to the emergence of fully modern speech, language and cognition appear to have occurred sometime in the period between 90,000 and 50,000 BP, the time frame between fossils like Skhul V and fully modern humans who were capable of talking as we do.

References Cited

ALBERT MA, FELDMAN RG, WILLIS AL. 1974. The "subcortical dementia" of progressive supranuclear palsy. *Journal of Neurology, Neurosurgery, and Psychiatry* 37:121-130.

ALDRIDGE J. W.., K. C. BERRIDGE, M. HERMAN AND L. ZIMMER. 1993. Neuronal coding of serial order: Syntax of grooming in the neostratum. *Psychological Science* 4:391-393.

ALEXANDER G. E., M. R. DELONG AND P. L STRICK. 1986. Parallel organization of segregated circuits linking basal ganglia and cortex. *Annual Revue of Neuroscience* 9:357-381.

ALEXANDER G. E. AND M. D. CRUTCHER. 1990. Functional architecture of basal ganglia circuits: Neural substitutes of parallel processing. *TINS* 13:266-271.

ALEXANDER M. P., M. A. NAESER AND C. L. PALUMBO. 1987. Correlations of subcortical CT lesion sites and aphasia profiles. *Brain*. 110:961-991.

BADDELEY A. D. 1986. *Working memory*, Oxford:Clarendon Press.

BAER, T., J. C. GORE, L. C. GRACCO AND P. W. NYE. 1991. Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels. *Journal of the Acoustical Society of America.* 90:799-828.

BAUM S. R., S. E. BLUMSTEIN M. A. NAESER AND C. L. PALUMBO. 1990. Temporal dimensions of consonant and vowel production: An acoustic and CT scan analysis of aphasic speech. *Brain and Language* 39:33-56.

BECKMAN, M. E., T-P. JUNG, S-H. LEE, S-H, K. DE JONG, A. K. KRISHNAMURTHY, S. C. AHALT, K. B. COHEN AND M. J. COLLINS. 1995. Variability in the production of quantal vowels revisited. *Journal of the Acoustical Society of America*. 97:471-489.

BENSON D. F. AND N. GESCHWIND. 1985. Aphasia and related disorders: A clinical approach. In *Principles of Behavioral Neurology* pp. 193-228, Editor Mesulam MM. Philadelphia: F. A. Davis.

BERRIDGE, K. C. AND I. Q. WHITSHAW. 1992. Cortex, striatum and cerebellum: Control of serial order in a grooming sequence. *Experimental Brain Research* 90:275-290.

BLUMSTEIN, S. E. 1994. The neurobiology of the sound structure of language. In M. S. Gazzaniga Ed. *The cognitive neurosciences* Cambridge Mass.:MIT Press.

BLUMSTEIN S. E. 1995. The neurobiology of language. In *Speech, Language and Communication* San Diego CA:Academic Press. p 339-370.

BLUMSTEIN S. E., W. E. COOPER, H. GOODGLASS, S. STATLENDER AND J. GOTTLIEB. 1980. Production deficits in aphasia: a voice-onset time analysis. *Brain and Language* 9:153-170.

BOE, L-J, S. MAEDA AND J-L HEIM. 1999. Neanderthal man was not morphologically handicapped for speech. *Evolution of Communication* 3:49-77.

BOE, L-J, J-L HEIM, K. HONDA, AND S. MAEDA. 2002. The potential Neanderthal vowel space was as large as that of modern humans. *Journal of Phonetics* 30:465-484.

BOSMA, J. F. 1975. Anatomic and physiologic development of the speech apparatus. In *Human communication and its disorders*, ed. D. B. Towers, 469-481. New York: Raven.

BOULE, M. 1911-1913. L'homme fossile de la Chapelle-aux-Saints. *Annales Paleontologie* 6:109; 7:21, 85; 8:1.

BRAMBLE, D. M. AND D. E. LIEBERMAN. 2004. Endurance running and the evolution of Homo. *Nature* 432:345-352.

BROCA P. 1861. Remarques sur le siege de la faculte de la parole articulee, suivies d'une observation d'aphemie (perte de parole). *Bulletin de la Societe d`Anatomie (Paris)* 36:330-357.

BUHR, R. D. 1980. The emergence of vowels in an infant. *Journal of Speech and Hearing Research* 23:75-94.

BURKE, R.E., S. O. FRANKLIN AND C. E. INTURRISI. 1994. Acute persistent suppression of preproenkephaline mRNA expression in the striatum following developmental hypoxic-ischemic injury. *Journal of Neurochemistry* 62, 1878-1886.

CARRE, R., B. LINDBLOM AND P. MACNEILAGE. 1995. Acoustic factors in the evolution of the human vocal tract. *C. R. Academie des Sciences Paris,* t 320, Serie IIb, 471-476.

CHIBA, T. AND J. KAJIYAMA. 1941. *The vowel: Its nature and structure*. Tokyo: Tokyo-Kaisekan Publishing Co.

CHIE, U., Y. INOUE, M. KIMURA, E. KIRINO, S. NAGAOKA, M. ABE, T. NAGATA, H. ARAI. 2004. Irreversible subcortical dementia following high altitude illness. *High Altitude Medicine and Biology* 5:77-81

CHOMSKY, N. 1995. *The minimalist program.* Cambridge MA: MIT Press.

COOLS R., R. A. BARKER, G. J. SAHAKIAN AND T. W. ROBBINS. 2001. Mechanisms of cognitive set flexibility in Parkinson's Disease. *Brain* 124:2503-2512.

CUMMINGS J. L. AND D. F. BENSON. 1984. Subcortical dementia: Review of an emerging concept. *Archives of Neurology* 41:874-879.

CUMMINGS J. L. 1993. Frontal-Subcortical circuits and human behavior. *Archives of Neurology* 50:873-880.

CUNNINGTON R., R. IANSEK, J. L. BRADSHAW AND J. G. PHILLIPS. 1995. Movement-related potentials in Parkinson's disease: Presence and predictability of temporal and spatial cues. *Brain.* 118:935-950.

DARWIN, C. 1859 *On the origin of species*: Facsimile ed. 1964 Cambridge Mass: Harvard University Press.

DELONG M. R. 1993. Overview of basal ganglia function. *In Role of the cerebellum and basal ganglia in voluntary movement.* Mano N, Hamada I, DeLong MR, editors. Amsterdam:Elsevier

D'ESPOSITO M. AND M. P. ALEXANDER. 1995. Subcortical Aphasia: Distinct profiles following left putaminal hemorrhage. *Neurology*, 45:38-41.

DOBZHANSKY, T. 1973. Nothing in biology makes sense except in the light of evolution.*American Biology Teacher.*35:125-129.

DRONKERS N. F., J. K. SHAPIRO, B. REDFERN AND R. T. KNIGHT. 1992 The role of Broca's area in Broca's aphasia. *Journal of Clinical and Experimental Neuropsychology* 14:session 8, Lang Aphasia.,

ENARD, W., M. PREZEWORSKI, S. E. FISHER, C. S. LAI, V. WIEBE, T. KATANO, A. P. MONACO,and S. PAABO. 2002. Molecular evolution of FOXP2, a gene involved in speech and language.*Nature.*41:869-872

FANT, G. 1960. *Acoustic theory of speech production*. The Hague: Mouton.

FEINBERG, M. J. AND O. EKBERG. 1990. Deglutition after near-fatal choking episode: radiologic evaluation. *Radiology* 176:637-640.

FISHER, S. E., F. VARGHA-KHADEM, K. E. WATKINS, A. P. MONACO AND M. E. PEMBREY, M. E. 1998. Localization of a gene implicated in a severe speech and language disorder*. Nature Genetics* 18:168-170.

FITCH W. T. 1997. Vocal tract length and formant frequency dispersion correlate with body size in macaque monkeys.*Journal of the Acoustical Society of America.* 102:1213-1222.

------ 2000a. Skull dimensions in relation to body size in nonhuman mammals: The causal bases for acoustic allometry. *Zoology* 103: 40-58.

------ 2000b. The evolution of speech: a comparative view. *Trends in Cognitive Science* 4:258-267.

FITCH, W. T. AND J. GIEDD. 1999. Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *Journal of the Acoustical Society of America* 106:1511-1522.

FITCH, W. T. AND D. REBY. 2001. The descended larynx is not uniquely human. *Proceedings of Royal Society London* B 268:1669-1675.

FLOWERS K.A AND C. ROBERTSON. 1985. The effects of Parkinson's disease on the ability to maintain a mental set*. J Neurology, Neurosurgery, Psychiatry* 48:517-529.

FUJIMURA, O AND Y. KAKITA, Y. 1979. Remarks on quantitative description of lingual articulation. In Lindbolm, B. and Ohman, S. (Eds) *Frontiers of Speech Communication Reserach.* Academic:London, pp. 17-24.

GEORGE, S. L. 1978. A longitudinal and cross-sectional analysis of the growth of the postnatal cranial base angle. *American Journal of Physical Anthropology* 49:171-178.

GOODALL J. 1986. *The chimpanzees of Gombe: Patterns of behavior*. Cambridge Mass: Harvard.

GOLDSTEIN K. 1948. *Language and language disturbances*. New York: Grune and Stratton.

GRAYBIEL A. M. 1995. Building action repertoires: memory and learning functions of the basal ganglia. *Current opinion in neurobiology* 5:733-741.

------- 1997. The basal ganglia and cognitive pattern generators. *Schizophrenia Bulletin* 23:459-469.

------- 1998. The basal ganglia and chunking of action repertoires. *Neurobiology memory learning* 70:119-136.

GRAYBIEL A. M, T. AOSAKI, A. W. FLAHERTY AND M. KIMURA. 1994. The basal ganglia and adaptive motor control. *Science* 265:1826-1831.

GREENBERG B. D., D. L MURPHY AND S. A. RASMUSSEN. 2000. Neuroanatomically based approaches to obsessive-compulsive disorder: Neurosurgery and transcranial magnetic stimulation. *The Psychiatric Clinics of North America* 23:671-685.

GREENBERG J. 1963*. Universals of language* Cambridge MA.: MIT Press.

GROSSMAN M. G., S. CARVELL. S. GOLLOMP, M. B. STERN, G. VERNON AND H. I. HURTIG. 1991. Sentence comprehension and praxis deficits in Parkinson's disease, *Neurology* 41:1620-1628.

GROSSMAN M. G., S. CARVELL, S. GOLLOMP, M. B. STERN, M. REIVICH, D. MORRISON, A. ALAVI AND H. L. HURTIG. 1993. Cognitive and physiological substrates of impaired sentence processing in Parkinson's Disease*. Journal of Cognitive Neuroscience* 5:480-498.

HARRINGTON D. L. AND L. HAALAND. 1991. Sequencing in Parkinson's Disease: Abnormalities in programming and controlling movement. *Brain* 114:99-115.

HAUSER, M. D., N. CHOMSKY AND W. T. FITCH. 2002. The faculty of language. What is it, who has it, and how did it evolve? *Science.* 298:1569-1579.

HEIM, J-L. 1989. La nouvelle reconstitution du crane neanderthalien de la Chapelle-aux-Saints. Methode et resultats. *Bulletin et Memoires de la Societe d'Anthropologie de Paris.* n. s., I:95-118.

HELLWAG, C. 1781. *De Formatione Loquelae*, Dissertation, Tubingen.

HENKE, W. L. 1966. Dynamic *articulatory model of speech production using computer simulation.* PhD dissertation, MIT.

HIIEMAE, K. M., J. B. PALMER, S. W. MEDICIS, J. HEGENER, B. S. JACKSON, AND D. E. LIEBERMAN. 2002. Hyoid and tongue movements in speaking and eating. *Archives of Oral Biology* 47: 11-27.

HILLENBRAND, J. L., A. GETTY, M. J. CLARK AND K. WHEELER. 1995. Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America,* 97:3099-3111.

HOCHSTADT, J. 2004. *The nature and causes of sentence comprehension deficits in Parkinson's disease: insights from eye tracking during sentence picture matching*. Ph.D disst. Brown University.

HOEHN M. M. AND M. D. YAHR. 1967. Parkinsonism: onset, progression and mortality. *Neurology*, 17:427-442

HONDA, K. AND M.K. TIEDE. 1998. An MRI study on the relationship between oral cavity shape and larynx position. *In Proceedings of the 5th International Conference on Spoken Language Processing* 2: 437-440.

HOWELLS, W. W. 1976. Neanderthal man: facts and figures. In *Proceedings of the Ninth International Congress of Anthropological and Ethnological Sciences*, Chicago 1973. The Hague: Mouton.

HOWELLS, W. W. 1989. *Skull Shapes and the Map; Craniometric Analyses in the Dispersion of Modern Homo.* Cambridge Mass: Papers of the Peabody Museum of Archaeology and Ethnology, Harvard University, Volume 79.

ILLES J., E. J. METTER, W. R. HANSON AND S. IRITANI. 1988. Language production in Parkinson's disease: Acoustic and Linguistic considerations. *Brain and Language* 33:146-160.

INOUE, T., H. KATO, T. ARAKI AND K. KOGURE 1992. Emphasised selective vulnerability after repeated nonlethal cerebral ischemic insults in rats. *Stroke* 23, 739-745.

IRWIN, O. C. 1948. Infant speech: development of vowel sounds. *Journal of Speech and Hearing Disorders* 13:31-34.

ISHIDA, R., J. B. PALMER, AND K. M. HIIEMAE. 2002. Hyoid motion during swallowing; factors affecting forward and upward displacement. *Dysphagia* 17:262-272.

JACKENDOFF, R. 1994. *Patterns in the mind: Language and human nature.* New York: Basic Books.

JELLINGER, K. 1990. New developments in the pathology of Parkinson's disease. In *Advances in Neurology. Vol. 53: Parkinson's Disease: Anatomy, Pathology and Theraphy,* Streifler MB, Korezyn AD, Melamed J, Youdim MBH, editors. New York: Raven Press. p 1-15

JEONG, J. H., J. C. KWON, J. H. CHIN, S. J. YOON, AND D. L. NA. 2002. Globus pallidus lesions associated with high mountain climbing. *Journal of Korean Medical Science* 17:861-863.

JUST, M. A., CARPENTER, P. A., KELLER, T. A., EDDY, W. F. M., AND THULBORN, K. R. 1996. Brain activation modulated by sentence comprehension. *Science* 274:114-116.

KIMURA, M., T. AOSAKI AND A. GRAYBIEL. 1993. Role of basal ganglia in the acquisition and initiation of learned movement. In Nano N, Hamada I, DeLong MR, editors. Role of the Cerrebellum and Basal Ganglia in Voluntary Movements, Amsterdam: Elsevier. p 83-87.

KLEIN D., R. J. ZATORRE, B. MILNER, E. MEYER AND A. C. EVANS. 1994. Left putaminal activation when speaking a second language; evidence from PET. NeuroReport 5:2295-2297.

KLEIN R. G. 1999. The human career, 2nd edition, Chicago:Chicago University Press.

KOTZ, S. A., M. MEYER, K. ALTER, M. BESSON, D. Y. VON CRAMON AND A. FREDERICI. 2003 On the lateralization of emotional prosody: An fMRI investigation. *Brain and Language,* 96: 366-376.

KRINGS, M., A. STONE, R. W. SCHMITZ, H. KRAINITZKI, M. STONEKING, AND S. PAABO. 1997. Neanderthal DNA sequences and the origin of modern humans. *Cell* 90:19-30.

KUHL, P. K., K. A. WILLIAMS, F. LACERDA, K. N. STEVENS, AND B. LINDBLOM. (1992) Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255:606-608.

LADEFOGED, P. J. AND BROADBENT, D. E. 1957. Information conveyed by vowels. *Journal of the Acosutical society of America* 29:98-104.

LADEFOGED, P., J. DE CLERK, M. LINDAU AND G. PAPCUN. 1972. An auditory-motor theory of speech production. *UCLA Working Papers in Phonetics* 22:48-76.

LAI, S. J., S. E. FISHER, J. A. HURST, F. VARGHA-KHADEM AND A. P. MONACO. 2001. A forkhead-domain dene is mutated in a severe speech and language disorder. *Nature* 413:519-523.

LAI, C. S, D. GERRELLI, A. P. MONACO, S. E. FISHER, AND A. J. COPP. 2003. FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder. *Brain* 126:2455-2462.

LAITMAN, J. T. AND E. S. CRELIN. 1976. Postnatal development of the basicranium and vocal tract region in man. In *Symposium on development of the basicranium* Ed. J. Bosma. 206-219. Washington D. C., : U.S. Government Printing Office.

LAITMAN, J. T., R. C. HEIMBUCH, AND E. S. CRELIN. 1979. The basicranium of fossil hominids as an indicator of their upper respiratory systems. *American Journal of physical Anthropology* 51:15-34.

LAITMAN, J. T. AND R. C. HEIMBUCH. 1982. The basicranium of Plio-Pleistocene hominids as an indicator of their upper respiratory systems. *American Journal of Physical Anthropology* 59:323-344.

LANGE, K. W., T. W. ROBBINS, C. D. MARSDEN, M. JAMES, A. M. OWEN AND G. M. PAUL. 1992. L-Dopa withdrawal in Parkinson's disease selectively impairs cognitive performance in tests sensitive to frontal lobe dysfunction. *Psychopharmacology* 107:394-404.

LAPLANE, D., M. BAULAC AND D. WIDLOCHER. 1984. Pure psychic akinesia with bilateral lesions of basal ganglia. *Journal of Neurology, Neurosurgery and Psychiatry* 47:377-385.

LAPLANE, D., M. LEVASSEUR, B. PILLON, R. BUBOIS, M. BAULAC, S. TRAN DINH, G. SETTE, F. DANZE AND J. C. BARON. 1989. Obsessive-Compulsive and other behavioral changes with bilateral basal ganglia lesions. *Brain* 112:699-725.

LASHLEY, K.S. 1951. The problem of serial order in behavior. In L. A. Jefress (Ed.) *Cerebral mechanisms in behavior* (pp. 112-146) New York:Wiley.

LEHERICY, S., M. DUCROS, P-F VAN DE MOORTELE, C. FRANCOIS, L. THIVARD, C. POOPON, N. SWINDALE, K. UGURBIL AND D-S KIM. 2004. Diffusion tensor tracking shows distinct corticostriatal circuits in humans. *Annals of Neurology* 55:522-529.

LIBERMAN, A. M., F. S. COOPER, D. P. SHANKWEILER, AND M. STUDDERT-KENNEDY (1967) Perception of the speech code. *Psychological Review* 74:431-461

LICHTHEIM, L. (1885) On aphasia *Brain* 7:433-484.

LIEBERMAN, D. E. 1995. Testing hypotheses about recent human evolution from skulls. *Current Anthropology* 36:159-198.

LIEBERMAN, D. E. AND R. C. MCCARTHY. 1999. The ontogeny of cranial base angulation in humans and chimpanzees and its implications for reconstructing pharyngeal dimensions. *Journal of Human Evolution.* 36:487-517.

LIEBERMAN, D. E., C. F. ROSS AND M. J. RAVOSA. 2000. The primate cranial base: Ontogeny, Function and Integration. *Yearbook of Physical Anthropology* 43:117-169.

LIEBERMAN, D. E., MCCARTHY, R.C., HIIEMAE, K. M. AND PALMER, J.B., 2001. Ontogeny of postnatal hyoid and laryngeal descent: implications for deglutition and vocalization. *Archives of Oral Biology* 46:117-128.

LIEBERMAN, P. 1968. Primate vocalizations and human linguistic ability. *Journal of the Acoustical Society of America* 44:1157-1164.

------ 1984. *The Biology and Evolution of Language*. Cambridge, Mass: Harvard University Press

------ 2000*. Human language and our reptilian brain: The subcortical bases of speech, syntax, and thought.* Cambridge Mass: Harvard University Press.

------ 2002. On the nature and evolution of the neural bases of human language. *Yearbook of Physical Anthropology*. 45:36-62

------ 2006. *Toward an evolutionary biology of language*. Cambridge Mass: Harvard University Press.

------- 2006. Limits on tongue deformation – Diana monkey vocalizations and the impossible vocal tract shapes proposed by Riede et al. (2005). *Journal of Human Evolution.*50:219-221.

------- in press . Current views on Neanderthal speech capabilities: A reply to Boe et al.,(2002. *Journal of Phonetics.*

LIEBERMAN, P., D. H. KLATT, AND W. H. WILSON. 1969. Vocal tract limitations on the vowel repertoires of rhesus monkey and other nonhuman primates. *Science* 164:1185-1187.

LIEBERMAN, P. AND E. S. CRELIN. 1971. On the speech of Neanderthal man. *Linguistic Inquiry*. 2:203-222.

LIEBERMAN, P., E. S. CRELIN AND D. H. KLATT. 1972. Phonetic ability and related anatomy of the newborn, adult human, Neanderthal man, and the chimpanzee. *American Anthropologist* 74:287-307.

LIEBERMAN P, E. T. KAKO, J. FRIEDMAN, G. TAJCHMAN, L. S. FELDMAN AND E. B. JIMINEZ. 1992. Speech production, syntax comprehension, and cognitive deficits in Parkinson's disease. *Brain and Language* 43:169-189.

LIEBERMAN P, B. G. KANKI, A. PROTOPAPAS, E. REED AND J. W. YOUNGS. 1994. Cognitive defects at altitude. *Nature* 372:325.

LIEBERMAN P., A. MOREY, J. HOCHSTADT, M. LARSON AND S. MATHER. 2005. Mount Everest: A space-analog for speech monitoring of cognitive deficits and stress. *Aviation, Space and Environmental Medicine*.76:198-207.

LIEGEOIS, F., T. BALDEWEG, A. CONNELLY, D. G. GADIAN, M. MISHKIN AND F. VARGHA-KHADEM. 2003. Language fMRI abnpormalitoies associated with FOXP2 gene mutation. *Nature Neuroscience* 6:1230-1237.

LISKER L. AND A. S. ABRAMSON. 1964. A cross language study of voicing in initial stops: acoustical measurements. *Word* 20:384-442.

LUBKER, J. AND T. GAY. 1982. Anticipatory labial coarticulation: Experimental, biological, and linguistic variables. *Journal of the Acoustical Society of America* 71:437-438.

MAHAJAN, P. V. AND B. A. BHARUCHA. 1994. Evaluation of short neck: Percentiles and linear correlations with height and sitting height. *Indian Pediatrics.* 31:1193-1203.

MARESCH, M.M. 1948. Growth of the heart related to bodily growth during childhood and adolescence. *Pediatrics* 2:382-402.

MARIN O, W. J. SMEETS AND A. GONZALEZ. 1998. Evolution of the basal ganlia in tetrapods: a new perspective based on recent studies in amphibians. *TNN* 21:487-494.

MARSDEN C. D AND J. A. OBESO. 1994. The functions of the basal ganglia and the paradox of sterotaxic surgery in Parkinson's disease. *Brain* 117:877-897.

MCCARTHY, R. C., D. S. STRAIT, F. YATES AND P. LIEBERMAN. forthcoming. The Origin of Human Speech.

MCCAMMON, R. 1952. *Human growth and development*. Thomas:Springfield

MCCOWAN, T. D. AND A. KEITH. 1939. The stone age of Mount Carmel, volume 2 of *The fossil remains from the Levalloiso-Mousterian.* New York: Clarendon Press

MIDDLETON F. A. AND P. L. STRICK. 1994. Anatomical evidence for cerebellar and basal ganglia involvement in higher cognition. *Science* 266:458-461.

MIRENOWICZ J AND W. SCHULTZ. 1996. Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. Nature 379:449-451.

MONCHI O., P. PETRIDES, V. PETRE, K. WORSLEY AND A. DAGHER. 2001. Wisconsin Card Sorting Revisited: Distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. *Journal of Neuroscience* 21:7733-7741.

MOUSE GENONE SEQUENCING CONSORTIUM. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420:520-562.

NAESER M.A., M. P. ALEXANDER, N. HELMS-ESTABROOKS, H. L. LEVINE, S. A. LAUGHLIN AND N. GESCHWIND N. Aphasia with predomininantly subcortical lesion sites; description of three capsular/putaminal aphasia syndromes. *Archives of Neurology* 39:2-14.

NATSOPOULOS D, G. GROUIOS, S. BOSTANTZOPOULOU, G. MENTENOPOULOS, Z. KATSAROU AND J. LOGOTHETIS. 1993. Algorithmic and heuristic strategies in comprehension of complement clauses by patients with Parkinson's Disease. *Neuropsychologia* 31:951-964.

NEAREY, T. 1978. *Phonetic features for vowels*. Bloomington: Indiana University Linguistics Club.

NEGUS, V. E. 1949.*The comparative anatomy and physiology of the larynx*. New York:Hafner.

NISHIMURA, T., A. MIKAMI, J. SUZUKI AND T. MATSUZAWA. 2003. Descent of the larynx in chimpanzee infants. *Proceedings of the National Academy of Sciences* 100:6930-6933.

OVCHINNIKOV, I. V., A. GOTHERSTROM, G. P. ROMANOVA, V. M. KHARITONOV, K. LIDEN, AND W. GOODWIN. 2000. Molecular analysis of Neanderthal DNA from the northern Caucasus. *Nature* 404:490-493.

PALMER, J. B., N. J. RUDIN, G. LARA AND A. W. CROMPTON. 1992. Coordination of mastication and swallowing. *Dysphagia* 7:187-200.

PARENT, A. 1986. *Comparative neurobiology of the basal ganglia* New York: John Wiley.

PETERSON, G. E AND H. L. BARNEY. 1952. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* 24:175-184.

PICKETT, E. R., E. KUNIHOLM, A. PROTOPAPAS, J. FRIEDMAN AND P. LIEBERMAN. 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* 36:173-188.

POLIT A. AND E. BIZZI. 1978. Processes controlling arm movements in monkeys. *Science* 201:1235-1237.

RENDALL, D., S. KOLLIAS, C. NEY AND P. LOYD. in press. Pitch (Fo) and formant profiles of human and vowel-like baboon grunts: The

role of vocalizer body size and voice-acoustic allometry. *The Journal of the Acoustical Society of America*

RIEDE, T., E. BRONSON, H. HATZIKIROU AND K. ZUBERBUHLER. 2005. Vocal production in a non-human primate; morphological data and a model. *Journal of Human Evolution.* 48:85-96.

RISSMAN, J., J. C. ELIASSEN AND S. E. BLUMSTEIN. 2003. Am event-related fMRI study of implicit semantic priming. *Journal of Cognitive Neuroscience.*

RUSSELL, G. O. 1928 *The Vowel*. Columbus:Ohio State University Press.

SANES, J.N, J. P. DONOGHUE, V. THANGARAJ, R. R. EDELMAN, AND S. WARACH. 1995. Shared neural substrates controlling hand movements in human motor cortex. *Science*. 268:1775-1777.

SCOTT, R. B., J. HARRISON, C. BOULTON, J. WILSON, R. GREGORY, S. PARKIN P. G. BAIN, C. JOINT, J. STEIN AND T. Z. AZIZ. 2002. Global attentional-executive sequelae following surgical lesions to globus pallidus interna. *Brain* 125:562-574.

SPURZHEIM, J. K. 1815.*The physiognomical system of Dr. Gall and Spurzheim.* London: Baldwin, Cradock and Joy.

STEVENS, K. N. 1972. Quantal nature of speech. In *Human communication: a unified view,* ed. E. E. David Jr., and P. B. Denes. New York: McGraw Hill, pp. 51-66.

STONE, M AND A. LUNDBERG. 1996. Three dimensional tongue surface shapes of English consonants and vowels. *Journal of the Acoustical Society of America* 99:3728-3736

STEVENS, K. N., AND A. S. HOUSE. 1955. Development of a quantitative description of vowel articulation. *Journal of the Acoustical Society of America* 27:484-493.

STORY, B. H., TITZE, I. R. AND HOFFMAN, E. A. 1996. Vocal tract area functions from magnetic resonance imaging*. Journal of the Acoustical Society of America*. 100:537-554.

STOWE, L. A., A. M-J. PAANS, A. A. WIJERS AND F. ZWARTS. 2004. Activation of "motor" and other non-language structures during sentence comprehension. Brain and Language 89:290-299.

STRINGER, C. B. 1998. Chronological and biogeographic perspectives on later human evolution. In T. Akazawa, K. Akoi, and O. Bar-Yosef (Editors) *Neanderthals and modern humans in western Asia*. New York: Plenum, pp. 29-38.

STRUB R. L. 1989. Frontal lobe syndrome in a patient with bilateral globus pallidus lesions. *Archives of Neurology* 46:1024-1027.

STUSS D. T AND D. F. BENSON. 1986. *The Frontal Lobes*. New York: Raven.

TAKAHASHI, K., F. C. LIU, K. HIROKAWA, AND H. TAKAHASHI. 2003. Expression of FoxP2, a gene involved in speech and language in the developing and adult striatum. *Journal of Neuroscience Research* 73:62-72.

THATCH, W. T. 1996. On the specific role of the cerebellum in motor learning and cognition: Clues from PET activation and lesion studies in man, *Behavioral and Brain Sciences* 19:411-431

TRUBY, H. L., J. F. BOSMA AND J. LIND. 1965. *Newborn infant cry*, Upsalla:Almquist and Wiksell.

VARGHA-KHADEM, F., K. WATKINS, R. PASSINGHAM AND P. FLETCHER. 1995 Cognitive and Praxic Deficits in a large family with a genetically transmitted speeech and language disorder. *Proceedings of the National Academy of Sciences* 95:2695-2700

VARGHA-KHADEM F., K. E. WATKINS, C. J. PRICE, J. ASHBRUNER, K. J. ALCOCK, A. CONNELLY, R. S. FRACKOWIAK, K. J. FRISTON, M. E.PEMBREY, M. MISHKIN, D. G. GADIAN AND R. E. PASSINGHAM. 1998. Neural basis of an inherited speech and language disorder. *Proceedings of the National Academy of Sciences* 95:2695-12700.

WATKINS, K. E., F. VARGHA-KHADEM, J ASHBURNER, R. E. PASSINGHAM, A. CONNELLY, K. J. FRISTON, R. S. J. FRACKIWIAK, M. MISKIN AND D. G. GADIAN. 2002. MRI analysis of an inherited speech and language disorder: structural brain abnormalities. *Brain* 125:465-478.

WEISENGRUBBER, G. E., G. FORSTENPOINTNER, G. PETERS, A. KUBBER-HEISS, AND W. T. FITCH. 2002. Hyoid apparatus and pharynx in the lion (Panthera leo), jaguar (Panthera onca), tiger (Panthera tigris), cheetah (Acinonyx jubatus) and domestic cat (Felis silvestris f. catus). *Journal of Anatomy* 201:195-201.

WERNICKE, C [1874] 1967. The aphasic symptom complex: A psychological study on a neurological basis. In *Proceedings of the Boston Colloquium for the Philosophy of Science*, vol 4. eds. R. S. Cohen and M. W. Wartofsky. Dordrecht:Reidel.

[i] This rules out the possibility of nonhuman primate airways such as those of Diana monkeys, being able to produce quantal vowels even if the degree of posterior pharyngeal expansion claimed by Riede et al. (2005)resulted in a 10:1 area function discontinuity because of its location (c.f. Lieberman, 2006)