

5-2013

## The Morphological Evolution of the Vocal Apparatus in the Human Lineage: Implications on the Emergence of Language

Jacob Jesch

Follow this and additional works at: [https://digitalcommons.lsu.edu/honors\\_etd](https://digitalcommons.lsu.edu/honors_etd)



Part of the [Anthropology Commons](#), and the [Geography Commons](#)

---

### Recommended Citation

Jesch, Jacob, "The Morphological Evolution of the Vocal Apparatus in the Human Lineage: Implications on the Emergence of Language" (2013). *Honors Theses*. 768.  
[https://digitalcommons.lsu.edu/honors\\_etd/768](https://digitalcommons.lsu.edu/honors_etd/768)

This Thesis is brought to you for free and open access by the Ogden Honors College at LSU Digital Commons. It has been accepted for inclusion in Honors Theses by an authorized administrator of LSU Digital Commons. For more information, please contact [ir@lsu.edu](mailto:ir@lsu.edu).

The Morphological Evolution of the Vocal Apparatus in the Human Lineage:  
Implications on the Emergence of Language

by

Jacob Jesch

Undergraduate honors thesis under the direction of

Dr. Mary J. Brody

Department of Geography & Anthropology

Submitted to the LSU Honors College in partial fulfillment of  
the Upper Division Honors Program.

May, 2013

Louisiana State University  
& Agricultural and Mechanical College  
Baton Rouge, Louisiana

## **Abstract**

Language is a defining characteristic of being human; however, its nature and origin remain elusive. One of the key aspects in the study of language is its biological framework, including the vocal apparatus. Accordingly, in order to approach language from this particular biological perspective, it is necessary to consider the morphological nature and evolutionary history of the human vocal tract. This paper chooses a descriptive and comparative approach to the subject, considering evidence from four different sources: vocal tract morphology and speech production in modern humans; differences between humans and nonhuman primates as well as other mammals in terms of vocal tract morphology and vocalization; maturation-derived differences between human infants and adults; and the emergence of humanlike vocal tract morphology and speech production as seen in the fossil record. Drawing on all these sources of evidence, the paper concludes that humanlike vocal tract morphology is essential for humanlike speech production and thus language, and is unique to *Homo sapiens* as a species.

## **Table of Contents**

Introduction	p. 4
Section I: Vocal Tract Morphology and Speech Production in Humans	p. 9
Section II: Vocal Tract Morphology and Speech Production in Nonhuman Primates	p. 33
Section III: Vocal Tract Morphology and Speech Production in Human Infants	p. 57
Section IV: Vocal Tract Morphology and Speech Production in Fossil Hominins	p. 64
Conclusion	p. 84
Acknowledgements	p. 87
Bibliography	p. 88

## Figures

Fig 4: The Three Subsystems of the Human Vocal Tract	p. 10
Fig. 2: The Hyo-Laryngeal Complex	p. 14
Fig 3: Cross-Sections of Vocal Tract Configuration, Cross-Sectional Area Functions, Acoustic Transfer Functions of F1 and F2, and Stylized Two-Tube Vocal Tract Configuration for the Vowels [i], [a], and [u]	p. 31
Fig 4: Two-tube Configuration and Permanently Descended Larynx in Humans, and Single-Tube Configuration and Non-Permanently Descended Larynx in Orangutans ( <i>Pongo borneus</i> ) and Capuchin Monkeys ( <i>Cebus</i> )	p. 36
Fig 5: Temporary Descent of the Larynx during Vocalization in Dogs and Primates	p. 42
Fig 6: Human Vowel Space in terms of F1 and F2, and Closest Chimpanzee Capability to Phonate [a], [i], and [u]	p. 49
Fig 7: Human Infant and Chimpanzee Single-Tube Vocal Tract, and Human Adult Bent Two-Tube Vocal Tract	p. 58
Fig 8: Crania and Reconstructed Laryngeal Structure and Reconstructed Vocal Tract for Modern Human Infants, Neanderthals, and Adult Modern Humans	p. 67
Fig 9: Hyoid Bone Morphology and Presence/Absence of hyoid bullae in gorillas, chimpanzees, australopithecines, Neanderthals, and modern humans	p. 71
Fig 10: Potential Maximum Vowel Space in Terms of F1 and F2 for an adult modern human male, an adult male Neanderthal, an adult modern human female or a 16-year old modern human male, a 10-year old modern human child, a 4-year old modern human child, and a newborn modern human infant	p. 75
Fig 11: Extent of the Temporalis Muscles in Western Lowland Gorillas ( <i>Gorilla gorilla</i> ) and modern humans	p. 79

## Introduction

Over the last few decades, the definition of being human has gradually changed, as more and more characteristics once thought to be uniquely human began to be observed in other extant animals as well as in the fossil record. Bipedalism definitively was established to have been the primary mode of locomotion of early australopithecines with the discovery of Lucy, a female *Australopithecus afarensis*, in 1974, and is now speculated to extend back in evolutionary time as far as seven million years ago (mya) to *Sahelanthropus tchadensis* (Conroy & Pontzer 2012). Additionally, bipedalism has been observed in extant hominoid species such as chimpanzees, bonobos, gorillas, and gibbons – albeit not habitual as it is in humans and their immediate ancestors. The presence of a large and well-developed brain compared to that of other extant and fossil primates has been known to modern science even before then with the discovery of Neanderthals in 1856 and their formal scientific description in 1864 by William King; that the trend toward such a development was not restricted to the genus *Homo* alone, but instead already emerging in earlier hominins was firmly established when Raymond Dart described and named *Australopithecus africanus* in 1925 (Conroy & Pontzer 2012). Tool use is also now known not to be exclusive to our species, having been observed in chimpanzees (Goodall 1986), gorillas (Breuer et al. 2005), and orangutans (Galdikas 1982), as well as numerous other primate and non-primate species. Likewise, the Oldowan industry – proposed by Louis Leakey in the 1930s (Leakey 1981) – gives evidence that tool use is not a recent innovation in the human lineage either.

Despite these substantial changes in perception of what makes humans human, one characteristic so far has been firmly upheld: language, which can be defined as the

communication of ideas and meaning via symbolic interaction. Humans are the only animals known to habitually engage in broad-scale communication using language. This uniqueness of human language as a form of vocalization is even more remarkable because it is such a complex effort of employing a system of encoded, abstract symbols, all for the sake of seemingly conveying trivialities most of the time (Desalles 2000) and not actually “useful” information. While the influence of “trivial” communication in terms of social bonding and the resulting long-term selective advantage stemming from a well-developed social group setting should not be entirely disregarded, most of human communication using language has no immediate impact on imminent survival.

While language itself is learned, the capacity as well as the drive to use it seems to be biologically innate to humans (Lenneberg 1967; Crain 1991). This assumption is supported by behavioral observations in human children. All infants are able to recognize human voices in general as well as specific types of vocalization as soon as twelve hours after birth (Mehler & Dupoux 1990; Aitchison 2000; Desalles 2000). Infants can pick up the differences in phonation between different human languages around four days of age (Mehler & Dupoux 1990). All children vocalize in the form of infant babbling beginning a few months after birth; even children born deaf do so until around six months of age (Lenneberg 1967). Finally, almost all children start using adult language, either spoken or signed, around one year of age (Aitchison 2000), after which they are able to pick up more than ten new words/meanings per day for a couple of years (Mehler & Dupoux 1990).

The natural mode of transmission for human language is speech, or vocalization (Fitch 2010), although human language is not exclusively restricted to it, as is evident in the

feasible use of sign languages; in fact, signed communication is often learned earlier by infants than spoken communication (Desalles 2000; Fitch 2010). Additionally, written communication is similarly effective as spoken communication in conveying information. Nevertheless, speech – and not sign – is the default transmission mode of the overwhelming majority of all human languages; similarly, all written languages are merely graphic representations of either spoken or – occasionally – signed language. For these reasons, speech’s fundamental importance for our species’ natural behavior points toward a biological adaptive predisposition for vocalized language use, requiring the ability to rapidly combine and/or rearrange many different sounds – phones – into words, phrases, and sentences at a sufficient rate for effective communication to occur (Nishimura et al. 2006). As such, human language must accommodate a vocal mode of transmission, and in fact it does so. There are no phonemes – the smallest units of meaning – in any human language that cannot be vocally expressed by humans; neither are there any phonemes that can be vocally expressed but not auditorily perceived (Lieberman 1975). Additionally, phones that are difficult to produce by a humanlike vocal apparatus are universally rare in human languages; at the same time, the universally most common phones are those that are easy to produce, sufficiently distinct from each other, as well as acoustically stable (Lieberman 1975). Lastly, humans are capable of producing as well as perceiving communicative vocalization at a much greater and effective rate than any other living primate species (Lieberman 1991).

This natural predisposition to spoken language is indicative of an evolutionary trend within the human lineage towards it, and the nature as well as time of emergence of language and a hypothetical preceding protolanguage have been debated by scientists



from many different fields over the last few centuries. Models of language evolution include a spontaneous, sudden appearance in evolutionary history once sufficient morphological, neural, and behavioral preconditions had been met (Chomsky 1972); a gradual emergence over time due to the retention of biological changes that proved to be evolutionarily advantageous, as according to a classical Darwinist model of natural selection (Lieberman 1975); and an extended oscillatory development of accelerated bursts of language emergence followed by decelerated periods of evolutionary stasis, which is closely inspired by the theory of evolutionary punctuated equilibria (Eldredge & Gould 1972). Some authors have argued that protolanguage arose as early as in australopithecines (Holloway 1983), others favor a time of appearance in early *Homo* (Falk 1980; Aiello & Dunbar 1993), and still others believe it arose as late as the Upper Paleolithic (Isaac 1976; Noble & Davidson 1991) and would thus have been exclusive to *Homo sapiens*. Additionally, language has been proposed to originate from earlier modes of primate vocalization (Steklis & Raleigh 1979; Dunbar 1993; Fitch & Hauser 1995) or alternatively a system of gestural communication (Hewes 1973; Calvin 1992; Armstrong et al. 1994). The latter theoretical perspective argues that neural prerequisites for fully developed language use arose earlier than morphological ones, and protolanguage thus relied on gestural communication using the hands instead of vocal communication (Aitchison 2000). Supporters point to the importance of gesturing in modern human communication as well as the full-fledged status of sign languages, while detractors argue that gestural communication would have been evolutionarily inefficient due to the fact that the hands are occupied during communication and cannot be used for other purposes (Aitchison 2000).

As is evident from the diversity of theoretical perspectives concerning the nature of language, no single factor in itself is likely to be entirely responsible for its evolution, and many different approaches such as ethology, psychology, physiology, anthropology, neurology, and morphology/anatomy need to be considered to attain a holistic model of language emergence (Lieberman 1975). However, this does not necessarily mean that all these approaches are tightly interconnected all the time; e.g., the fact that vocalization and language are separate biological concepts has been known since 1861, when Pierre Paul Broca discovered that damage to an area of the brain that has since been named “Broca’s area” in his honor would prevent affected individuals from articulating coherently, reading, and writing, but not from actual phonation such as singing (Broca 1861). As such, this literary review focuses on one particular approach to language, namely the role of the morphological/anatomical nature of the human vocal apparatus in speech production, as well as its evolutionary history. The paper is divided into four thematically-organized sections. In section one, I first discuss the nature of modern human language, both acoustically and articulatorily, and its morphological basis. Then, in section two, I take a look at the vocalization of nonhuman primates, focusing on a comparative approach concerning the types of vocalizations they naturally employ as well as their capabilities of emulating humanlike language. I also consider some evidence from non-primate animals, such as parrots. In section three, I then take a look at the anatomical differences of the vocal apparatus between human infants and adults, as well as the vocal tract’s maturation process and the changes accompanying it. Lastly, in section four, I turn to the fossil record, comparing it to the morphology of adult modern humans and infants, as well as nonhuman primates. Based on this evidence, I draw some

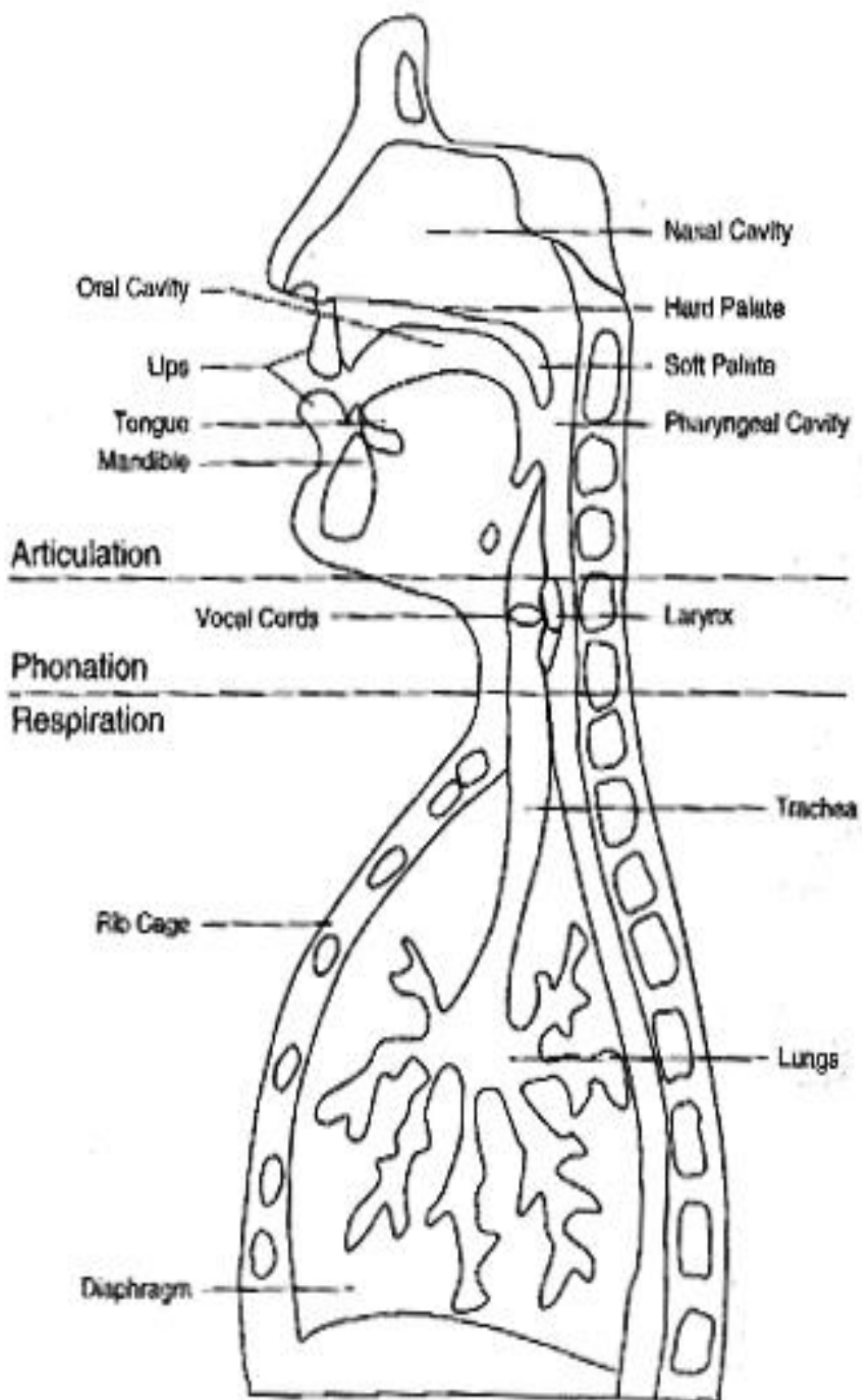
implications concerning language capability and its emergence in the evolutionary development of the human lineage.

### **Vocal Tract Morphology and Speech Production in Humans**

The human vocal tract is traditionally subdivided into three different parts or subsystems: respiratory, phonatory, and articulatory subsystem (MacNeilage 2008; Fig. 1). Each of these subsystems has its own distinctive role in the vocalization process and all can operate independently from each other, which contributes to the relatively great variety and rate of human vocalization (Desalles 2000). In addition, each subsystem also has a semi-independent evolutionary history, and while all three are related to each other by function and origin, their modern forms did not appear within the same evolutionary time frame.

The respiratory subsystem includes the lungs, the diaphragm, the trachea, and the rib cage. It is the oldest of the three subsystems in that its present-day configuration has remained relatively unchanged for the longest period of evolutionary time compared to the other two subsystems (Shubin 2008). Evidence for the presence of respiratory organs similar to modern ones has been found in some of the earliest tetrapod fossils and predates the amphibian-reptile divergence point (Shubin 2008); the heavy reliance on communicative vocalization found in some taxa on either phylogenetic branch – e.g., many species of frogs, and songbirds as well as humans – supports this assumption (Fitch 2010).

The main purpose of the respiratory subsystem, besides its other applications such as respiration, is to provide a source of acoustic energy for vocalization (Lieberman 1975;



*Fig. 1: The Three Subsystems of the Human Vocal Tract; MacNeilage (2008)*

Lieberman 1991; Nishimura et al. 2006; MacNeilage 2008; Ghazanfar & Rendall 2008; Fitch 2010). This energy source consists of air flow that exits the vocal tract during expiration, with the volume of air taken in during inspiration being directly proportional to the intended length of phonation that follows (Lieberman 1991). Elastic recoil subsequently causes the lungs to deflate during expiration, with subglottal pressure being maintained through muscle action by the internal intercostal and abdominal muscles (Draper et al. 1959; Hixon et al. 1976; MacNeilage 2008). From a physiological perspective, this source of acoustic energy is “free” in that it would be present regardless of whether vocalization occurs or not due to the fact that respiration is required for survival; the vocalization effort is thus just a small additional energy output added to this process (Moon & Lindblom 2003; Fitch 2010). While the overwhelming majority of vocalization makes use of this energy source, there are a few instances of vocalization in which the source of energy instead comes from inspiratory air flow (e.g., laughter, infant cries) or the creation of a vacuum in the lungs (e.g., click sounds, which are employed in some Khoisan languages of Southwest Africa) (Fitch 2010). With the exception of these few cases, vocalization generally takes place during expiration and modifies the process accordingly; while a normal expiration phase lasts ~2s, during vocalization restrictions on elastic control of the lungs prevent deflation, thereby extending the expiration phase up to ~40s and thus providing both more time as well as increased stability for speech production (Lieberman 1975).

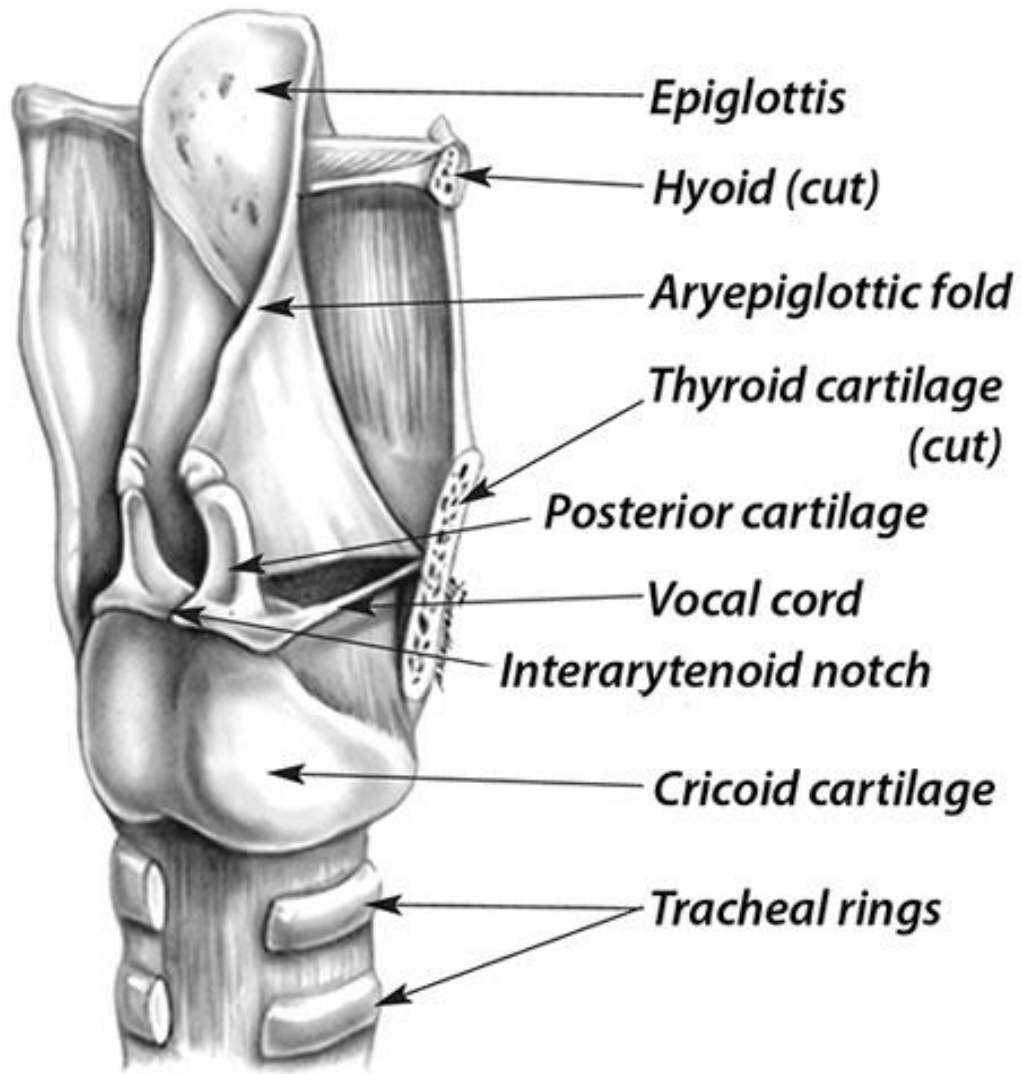
In most instances of human vocalization, air flow is organized into breath groups, or “relatively steady patterns of airflow that are the result of coordinated respiratory maneuvers” (Lieberman 1975:42) which are maintained during speech. For example, a

particular utterance, such as a sentence that is preceded and followed by an inspiratory pause, constitutes one breath group (Lieberman 1975; Ghazanfar & Rendall 2008). Breath groups are critical linguistic features of tone languages, and carry a significant degree of linguistic importance in almost all human languages (Lieberman 1975). To illustrate this, in many languages a fall in pitch during a breath group's duration – which is reflective of falling alveolar air pressure in preparation for inspiration – indicates a non-question statement, whereas a rise in pitch – which reflects a lessened dependence on air pressure for phonation due to muscular tension of the larynx – indicates a question (Lieberman 1975). The immense importance of breath groups for human language additionally indicates an early evolutionary emergence of the respiratory subsystem, and breath groups themselves or an equivalent phonetic feature were likely already present in any hypothetical form of speech-based protolanguage (Lieberman 1975).

A critical factor in human speech vocalization is the ability to fine-control respiration to a much greater extent than in any other primate, which allows for precise control of the subglottic air pressure that is the energy source for vocalization (MacLarnon & Hewitt 1999). Respiratory fine-control is determined by the cross-sectional area of the thoracic vertebral canal, which is determined by measuring the dorso-ventral diameter (narrowest height) and transverse diameter (narrowest width) of the vertebral canal through each thoracic vertebra (MacLarnon & Hewitt 1999). In primates, thoracic vertebral canal size is usually proportionate to body size and is thus often also a sexually dimorphic feature due to differences in body size between males and females in most taxa; human males and females fall on the upper extreme end of the range of deviation in primates, with males having slightly larger cross-sectional areas than females (MacLarnon & Hewitt

1999). Additionally, human thoracic vertebral innervation – the supply of nerves – is greatly enhanced compared to other primates, whereas the innervation for the rest of the spinal cord follows expected primate trends (MacLarnon 1987; MacLarnon 1993). Respiratory fine-control allows for extended vocalization through longer breath groups without the need to inject pauses for inspiration (Draper et al. 1959; Campbell 1968; Proctor 1974; Hixon & Weismer 1995; Ghazanfar & Rendall 2008) as well as modulation of intensity/loudness, emphasis on different parts of the same utterance, and pitch/intonation patterns (Ladefoged 1968; Proctor 1974; Hoit et al. 1990). While the exact degree of respiratory fine-control of subglottal air pressure differs individually based on lung volume (Mead et al. 1968) and the linguistic character of a given vocalization (Ladefoged 1968), as well as the relative importance of respiratory vs. laryngeal control for speech production (Ladefoged 1968; Hoit et al. 1990; Stathopoulos & Sapienza 1993), any given human individual's ability of respiratory fine-control would more than likely trump that of any other individual nonhuman primate, regardless of the nature of vocalization. Additionally, intensity and emphasis of speech rely primarily on respiratory fine-control (Lieberman 1984; Stathopoulos & Sapienza 1993), as does the production of some human languages' consonants (Ladefoged 1968).

The phonatory subsystem consists of the hyo-laryngeal complex, which includes the vocal cords, the larynx, the epiglottis, and the hyoid bone (Fig. 2). Although the hyoid bone is not formally a component of the vocal tract, it is usually included in this subsystem as well, due to its great importance in vocalization. The phonatory subsystem is structurally supported by the basicranium, together with the superiorly-situated articulatory subsystem (Fitch 2010). This subsystem's importance for speech production



*Fig. 2: The Hyo-Laryngeal Complex and Its Associated Components; after Gray (1918)*

has been known to science at least since the 17<sup>th</sup> century, when the similarity between human and nonhuman primate laryngeal morphology was noted (Lieberman 1975), and until the mid-20<sup>th</sup> century the general consensus in language studies was that the larynx is the most important component in speech production. Since then, this role has been reassigned to the supralaryngeal vocal tract (i.e., the articulatory subsystem), but the phonatory subsystem still remains critically important for vocalization.



Each of the components of the phonatory subsystem occupies a different role in the production of speech. The primary role of the vocal cords is filtering the air flow coming from the respiratory subsystem. This is accomplished by very rapid vibration of the vocal cords, emitting puffs of air in the process (Lieberman 1975). Vibration is initiated by an inward flexion of the vocal cords from their relaxed position due to laryngeal muscle adjustment. The vocal cords are then pushed open by expiratory air flow from the lung, which creates a force known as the “Bernoulli effect” that causes the vocal cords to be pushed back together again. Once the vocal cords are shut, air flow ceases, and static air pressure pushes the vocal cords open once more, causing the process to repeat. According to the myoelastic-aerodynamic theory of phonation, the vibration is a passive – or automatized – process in that it does not require continuous neural stimulation once set in motion, which greatly increases possible vibration rates (van den Berg 1958; Titze 1994). The exact rate of vocal cord vibration, which itself is influenced by vocal cord length and mass (Ghazanfar & Rendall 2008), determines the fundamental frequency of phonation, or  $F_0$ , which is measured in Hz (Lieberman 1975; Lieberman 1991; Fitch 2010). The magnitude of  $F_0$  determines a vocalization’s pitch, and is thus a critical component of breath groups (Lieberman 1975).

Naturally, fundamental frequencies are not restricted to human vocalization, but occur in the vocalization of all animals with vocal cords (Lieberman 1975); however, they are usually less important in these animals’ vocalization than is the case for humans. In humans as well as other animals, vocal cord vibration rate is adjustable depending on their shape/mass, the tension exerted from the laryngeal muscles, and lung-generated air pressure (Lieberman 1975). All these factors are influenced by age as well as sexual

dimorphism; in adult human males, F0 ranges from ~80-300 Hz, with the average being 125 Hz, whereas in adult human females and children, it can range up to 500 Hz, with the average being 225 Hz for adult females and 350 Hz for infants (Lieberman 1975; MacNeilage 2008). This discrepancy is caused by the greater length of adult male vocal cords compared to those of adult females and children (Lieberman 1975). While F0 tends to fall progressively over the course of one breath group in most animals, including most nonhuman primates (Andrew 1963; Hauser 1991), humans can control it voluntarily and thus can maintain steady breath groups; some nonhuman primates that have very well-developed laryngeal air sacs can also do this, but not as efficiently or sophisticatedly as humans (MacLarnon & Hewitt 1999).

Whether a particular utterance is voiced or voiceless depends on vocal cord action. Voiced speech is a result of the above-described vibratory patterns and requires some preconfiguration of both the phonatory and articulatory subsystem through muscle action by the lateral cricoarytenoid, internal and external thyroarytenoid, and interarytenoid muscles (Lieberman 1975). On the other hand, voiceless speech – e.g., whispering – is caused by turbulences of air being rushed through constricted vocal cords; any further vocal modification of voiceless speech is entirely dependent on the supralaryngeal vocal tract (Lieberman 1975). Humans are able to abduct their vocal cords, meaning they can briefly draw them out of the air flow, creating voiceless sounds amidst a stream of voiced vocalization in the process (MacNeilage 2008). This ability gives humans almost twice the theoretical consonant and voiceless vowel repertoire compared to other animals unable to do so (MacNeilage 2008). Additionally, vocal cord length in humans as well as other primates is relatively long compared to other animals that have shorter vocal cords

for more efficient respiration, and is accompanied by shortened arytenoid cartilages which restrict vocal cord expansion, the combination of which eases vibration for phonation purposes (Negus 1949).

The vocal cords are housed in the larynx, sometimes called “voice box”. The original purpose of the larynx in evolutionary terms seems to have been as a protective lid covering the trachea in order to shield the lungs (Lieberman 1975). In many mammals, including all primates, laryngeal adaptations have occurred that sacrifice respiratory efficiency for enhanced phonation (Negus 1949), and this especially has been the case in humans, whose larynx is only half the diameter of the trachea. While the larynx is not the primary source of human speech capabilities as was once believed, its importance rests in its function of housing the vocal cords and thus serving as the main filter through which acoustic energy passes (Lieberman 1975). Thus, individuals that lack a larynx would still be able to communicate, but only via voiceless speech, i.e., whispering (Lieberman 1975).

One of the defining features of the human vocal tract is the permanent descent of the larynx, which occurs during maturation. In adult humans, the larynx is situated posterior-inferiorly in relation to the mandible and is adjacent to the boundary of the fifth and sixth cervical vertebrae, which is much lower than in a number of other mammals –e.g., cats’ and pigs’ larynges are adjacent to the third cervical vertebra, dogs’ larynges are adjacent to the second cervical vertebra, and horses’ larynges are adjacent to the first cervical vertebra (Barone 1976). However, while laryngeal descent has long been believed to be a uniquely human characteristic (Negus 1949; Lieberman 1984; Crelin 1987; Houghton 1993; Laitman & Reidenberg 1993), recent evidence showing permanent laryngeal

descent in other mammal species has revealed that this is not the case (Fitch & Reby 2001; Fitch 2002; Nishimura 2003; Nishimura et al. 2003; Nishimura 2005; Nishimura et al. 2006). Nevertheless, a descended larynx remains important for humanlike speech production; while computer modeling has shown that similarly sufficient articulatory abilities are possible for a moderate range of laryngeal heights (de Boer 2010) and are thus likely more dependent on precise articulatory control (Boe et al. 2002; Boe et al. 2007), a laryngeal height that completely deviates from this range would make humanlike vocalization impossible (Carre 2004; 2009); at the same time, the largest possible acoustic area is achieved for a laryngeal depth corresponding to a ~1:1 ratio of oral cavity and pharyngeal cavity length (Lieberman 1975; Lieberman et al. 1992), which is the natural state in adult humans (de Boer 2010). Laryngeal descent is directly responsible for the separation of the hyoid and cricoid cartilages (Lieberman 1975), and because of its lower position relative to the supralaryngeal vocal tract, the human larynx in general has much fewer cartilaginous appendages than most nonhuman primates' larynges (Aitchison 2000).

Laryngeal descent had some potentially wide-ranging morphological effects and might indirectly have contributed to the strong orthognathism and globular cranial shape found in humans. A descended larynx significantly reduces available basicranial space; thus, a globular cranial shape consequently becomes necessary to retain enough space for the brain (DuBrul 1958; Lieberman 1975; Aiello 1996). Additionally, the descent of the larynx could have enabled an enlargement of the tongue base (Aitchison 2000; Fitch 2010), which made possible the rounded shape and 90° bent configuration of the human tongue that is a prerequisite for modern human speech capabilities (Lieberman 1991;

Fitch 2010). In the latter case, however, the causal relationship is not clearly established; it is certainly possible that tongue base enlargement set in motion laryngeal descent instead. Also, no clear correlation has been established between laryngeal descent and basicranial angle in humans (Fitch & Giedd 1999), which is also representative of a lack of such correlation in mammals in general. The exact impact of laryngeal descent on orthognathism also remains somewhat spurious, as there is no observable laryngeal descent in certain very orthognathic breeds of toy dogs and cats and – perhaps more importantly – newborn humans (Fitch 2010).

Compared to other primates as well as in relationship to general sexual dimorphism in humans, human laryngeal sexual dimorphism is extreme. In males, laryngeal descent during puberty is 2-3cm greater than in females and vocal cord vibration is twice as long (Titze 1989; de Boer 2010), which results in the aforementioned lower F0 of adult males compared to adult females and children, and consequently the resulting deeper pitch of male voices. The male larynx also tends to be much more voluminous than the female larynx, although both sexes' larynges are relatively more voluminous than those of other mammals (Desalles 2000). In fact, computer modeling studies have suggested that male laryngeal descent actually somewhat compromises articulatory capability slightly, whereas female laryngeal descent puts the larynx at the optimal height for articulation purposes (de Boer 2010). These results are supported by studies that have found a greater vowel range in females than in males (Fant 1975; Diehl et al. 1996; de Boer 2009). A possible explanation for this obvious disadvantage is that in males, laryngeal position is a compromise between sufficient articulatory ability and size exaggeration purposes (de Boer 2010), which will be discussed in depth in section two.

The descent of the larynx in humans has also led to a number of adaptive disadvantages, the greatest of which is the increased risk of asphyxiation due to choking during swallowing. The main reason for this is that unlike many other mammals, humans are unable to connect their larynx to the velum and thus cannot seal their trachea during swallowing, which allows for the potentially hazardous intrusion of food particles (Lieberman 1991; de Boer 2010). Other disadvantages include lowered chewing efficiency, reduced olfactory capabilities due to an inability to seal the oral cavity during nasal inspiration, reduced respiratory efficiency, a higher risk of dental impaction and subsequent infection due to compaction of the mandible, and an inability to breathe while the oral cavity is filled with a liquid, which most other mammals can do (Negus 1949; Lieberman 1975; Lieberman 1991). The retention of such a feature such as a permanently descended larynx, despite the numerous adaptive disadvantages that accompany it, suggests that significant communicative or otherwise beneficial advantages offset the disadvantages in the evolutionary balance (Negus 1949; Lieberman 1975; Fitch 2010). Also, a few evolutionary innovations have appeared in humans that negate these disadvantages to some degree; e.g., the ability to temporarily close the vocal cords and contract the aryepiglottic folds through elevation of the hyoid bone, which pushes the epiglottis against the tongue base and causes the aryepiglottic folds to cover the laryngeal vestibule (Negus 1949) with the purpose of decreasing choking risk – a process that is nearly as complicated as it sounds and is not nearly as efficient as the simple ability of other animals to connect their larynx to the velum.

The hyoid bone, which is included in this discussion of the phonatory subsystem due to its relevance for the subsystem's function, serves as a muscle attachment site anchor

for the intrinsic muscles of the tongue as well as most other muscles of the vocal tract (Fitch 2010). In primates and also big cats of the genus *Panthera*, the epihyal (upper) portion of the hyoid bone is made from ligament and muscle, whereas the basihyoid (lower) portion is made from solid bone (Fitch 2010). The hyoid bone is a sesamoid bone, which means that it does not articulate directly with any other bone; instead, the hyoid bone connects to the rest of the skeleton via cartilages. Similarly to the larynx, in humans there is also a descent of the hyoid bone, although it is less extreme in comparison (Negus 1949; Lieberman 1984; Crelin 1987; Houghton 1993; Laitman & Reidenberg 1993). Also, as is the case for the larynx, the hyoid bone's importance for speech production has long been overestimated, especially in the fossil record – for apparent reasons of being a bone and thus subject to better preservation than the soft-tissued vocal tract (Boe et al. 2002). Of course, this is not to say that the hyoid bone has no significance in vocalization at all; on the contrary, critical muscle control would be impossible without a proper hyoid bone (Arensburg et al. 1989; Fitch 2010). Just as the larynx, the hyoid bone is also sexually dimorphic in humans, projecting more anteriorly in males due to differential growth of thyroid cartilage, which causes the prominent appearance of the Adam's apple in adult males (Falk 1975). Additionally, the position of the hyoid bone also correlates with age – the older the individual, the lower the position of the hyoid bone (Falk 1975).

The third subsystem, which is called the articulatory subsystem and includes the pharyngeal cavity and the oral cavity as well as the anatomical features associated with them – pharynx, palatine uvula, soft (velum) and hard palate, alveolar ridge, tongue, mandible and maxilla, teeth, and lips – is nowadays recognized as the primary component

of speech production in humans (Lieberman 1975). Due to its anterior position in relation to the larynx, the articulatory subsystem is also called the supralaryngeal vocal tract. In humans, it is radically different from other animals, as its primary purpose has shifted from being devoted exclusively to breathing, chewing, and swallowing to participating in speech production (Lieberman 1991). As such, features such as evenly sized teeth, interlaced facial muscles, and a muscular and flexible tongue all maintain other purposes than vocalization, but are also critically important for the latter because they make the vocalization process more stable and thus more efficient (Aitchison 2000).

The critical importance of the supralaryngeal vocal tract over that of the larynx for human speech production began to be fully understood around the middle of the 20<sup>th</sup> century with the refinement of the source-filter theory of speech production (Stevens & House 1955; Chiba & Kajiyama 1958; Fant 1960), which originally had been proposed more than one-hundred years earlier (Müller 1848). The source-filter theory of speech production establishes the role of the supralaryngeal vocal tract as an adjustable acoustic filter – i.e., it assumes an acoustic transfer function – that allows for transmittance of certain acoustic wavelengths while blocking out others, which is accomplished through continuous articulatory reconfiguration modifying both supralaryngeal vocal tract shape and length (Lieberman 1975). Because  $\text{wavelength} = \frac{\text{propagation velocity}}{\text{frequency}}$ , with propagation velocity being constant at normal atmospheric pressure (i.e., 1 atm), the transfer function of the supralaryngeal vocal tract filter is described in terms of frequency (Lieberman 1975). The initial frequency component of F0, which is determined by vocal cord vibration, is accompanied by additional frequency components at multiples of F0, which are called harmonics (Fitch 2010) – e.g., an F0 of 100 Hz would have harmonics at



200 Hz, 300 Hz, 400 Hz, and so on. In human vocalization, the modulation of the first 35 harmonics, ranging from 20 Hz to 15 kHz, constitutes the main acoustic source of phonation, as all other frequencies are absorbed/attenuated by the vocal tract walls (Lieberman 1975; Ghazanfar & Rendall 2008; MacNeilage 2008); out of these 35, only six harmonics, ranging up to 7 kHz, play a significant role in auditory speech perception, and only the lower three – F1, F2, and F3 – are required for phonetic distinction (Lieberman 1975; Fitch 2002) – with F3 only being required for the distinction of some consonants, such as [l] and [r] (de Boer 2010). These three harmonics, at frequencies of below 3 kHz (de Boer 2010) but not sequential or contiguous – i.e., F2 is not the immediate multiple of F1, and F3 is not the immediate multiple of F2 (Ladefoged & Johnson 2010) – and perceived acoustically as timbre (Ghazanfar & Rendall 2008), are called formant frequencies.

Formant frequencies are perhaps the most important of all acoustic features in human vocalization. Maximum acoustic energy passes through the supralaryngeal vocal tract filtering system at these frequencies, and the controlled variation thereof – i.e., which tone is dominant – is a defining factor of human speech compared to the vocalization of other primates (Fant 1960; Lieberman 1975; Lieberman 1991; Titze 1994; Fitch & Hauser 2002; Riede et al. 2005; Fitch 2010). The exact type of produced formant frequency is determined by the supralaryngeal vocal tract's cross-sectional area function, which is the sum of the volume of its shape and size (Lieberman 1975; Ghazanfar & Rendall 2008). The greatest factor in determining cross-sectional area is articulation of the mandible and the tongue body, which are responsible for the primary modulation of F1 and F2, respectively (de Boer 2010). Cross-sectional area function is inversely

proportional to formant frequency magnitude – e.g., a long vocal tract corresponds to lower formant frequencies – meaning that, on average, adult males tend to produce lower formant frequencies than adult females or children (Fant 1960; Fitch & Giedd 1999; Fitch 2002). In addition to the impact of the supralaryngeal vocal tract, formant frequencies are also secondarily influenced by laryngeal height, with higher-than-average positions causing a large reduction in F2 and lower-than-average positions causing a moderate reduction in F1 and F2 (de Boer 2010). Because each individual's vocal tract produces slightly different formant frequencies – hence why each individual's voice sounds different – they are also used for identification purposes in many species (Fitch 2002).

Due to the great flexibility of their supralaryngeal vocal tract, especially the tongue because of its internal musculature which makes it more flexible, articulate and manipulatable than that of any other primate (Aitchison 2000; Takemoto 2001), humans can very quickly produce changes in formant frequency during very short periods of vocalization by means of laryngeal contraction/expansion and subsequent articulatory modification (Lieberman 1975; Fitch 2010). A humanlike “bent” vocal tract –i.e., one in which the horizontal supralaryngeal vocal tract (oral cavity) and the vertical supralaryngeal vocal tract (pharyngeal cavity) are oriented at a 90°-angle with respect to each other – is not necessary for such flexibility, but makes it much easier to achieve (de Boer 2010). Additionally, a bent vocal tract allows for articulatory modifications to occur semi-independently from each other in the oral cavity as well as the pharyngeal cavity, respectively (Lieberman et al. 1969; Lieberman 1984; Fitch 2000a; Fitch & Hauser 2002). This precise and also voluntary control of the supralaryngeal vocal tract allows for a rapid-fire transmission of up to 15 phonemes/second (Deacon 1997), and consequently

vast amounts of potentially useful communication of information in terms of selective fitness (Lieberman 1991). At the same time, an accompanying critical fact is that humans are also able to actually perceive and process information at this rapid rate via language, with maximum sound perception for language being 5-9 segments/second – corresponding to ~15 phonemes/second (MacNeilage 2008) – and maximum speech unit perception (e.g., a syllable) being 20-30 segments/second (Lieberman 1975). Together, these capabilities constitute a rate of information processing that for humans has so far been impossible to achieve by any other means, such as gesturing (Lieberman 1975; Lieberman 1991). This overwhelming reliance on formant frequency modulation in humans is indicative of our primate heritage and its corresponding neural preadaptation; other animals that make extensive use of vocalization often employ a different mechanism, e.g. tonal transitioning in song birds (Lieberman 1991).

In order to accomplish the complex configurations required for an acoustic transfer function, the supralaryngeal vocal tract houses most of the repertoire of articulation points – places responsible for the formation of formant frequencies and thus specific phones – which are produced through vocal tract articulation by orofacial muscular tension. Articulation points are well-developed in most mammals (Fitch 2010) and are involved in the swallowing and lactation reflex (Wall & Smith 2001); in humans, they were then modified for speech production within the constraints of not marginalizing their original purposes (MacNeilage 1998). Human articulation points include labial (lips), interdental/dental (teeth), alveolar (alveolar ridge), palatal (hard palate), velar (velum), uvular (palatine uvula), pharyngeal (pharynx wall), and glottal (epiglottis/larynx), the last of which is the only major articulation point not associated

with the supralaryngeal vocal tract (MacNeilage 2008). Articulation point position is directly related to F1 and F2, with F1 being higher if the articulation point is more posteriorly and/or ventrally located, and F2 being lower if the articulation point is more posteriorly located (MacNeilage 2008; Ladefoged & Johnson 2010). Likewise, the supralaryngeal vocal tract is the anatomical device that enables articulation manners, or ways through which specific phones are produced. The most common articulation manners include stops, nasals, fricatives, affricates, liquids, and semi-vowels or glides (MacNeilage 2008). Most articulation manners can be combined with most articulation points to create a large selection of differential phones; e.g., in English, stops can be produced labially (voiced [b] and voiceless [p]), dentally (voiced [d] and voiceless [t]), palatately (voiced [g] and voiceless [k]), or glottally through shutting the vocal cords (Lieberman 1975).

Because a great amount of anatomical maneuvering is necessary for the rapid rate of vocalization humans engage in during normal speech production, supralaryngeal vocal tract reconfiguration happens in a constant mode of rearrangement in that it begins to adjust to the position required for the next upcoming phone before having fully assumed the position of the previous one (Lieberman 1991). This “liquidity” is called anticipatory coarticulation and results in a characteristic blending of phones within a word, phrase, or sentence, which is responsible both for the tendency to omit pauses between independent phonemic constructs such as words in natural speech, as well as that of certain allophones – e.g., the allophonic interchangeability of [s] and [z] in the English word “observe” (Lieberman 1975; Desalles 2000). From an evolutionary point of view, it could be argued that this was an adaptive response to the greater perception rate of speech units, which is

more than thrice that of individual sound perception – although the opposite could also be maintained. That anticipatory coarticulation and speech unit perception rate are closely related features is supported by the fact that early attempts to provide visually or otherwise impaired individuals with machines that had the purpose of reading out loud information to them failed miserably due to the machine's inability to blend in different phonemes in the time-varying patterns of formant frequencies that occur in natural human speech. Instead, this resulted in very unnatural-sounding speech that was both excruciatingly slow as well as generally incomprehensible (Lieberman 1975; Remez et al. 1981; Lieberman 1991). While the basis of anticipatory coarticulation is likely biologically innate due to its fundamental importance for natural human speech production, the exact nature of anticipatory coarticulation varies for speakers of different languages, dialects, and accents; it is thus a learned automatized behavior (Lieberman 1991).

A concept that is related to anticipatory coarticulation is phonation onset, or the delay or anticipation of vocal tract configuration during vocalization (Lieberman 1975). Whether phonation onset is delayed or anticipatory, and the exact amount of time for each, depends on the nature of the produced phonation - e.g., in English, the voiceless labial stop [p] normally has a delayed phonation onset of ~100ms, meaning that acoustic energy exits the supralaryngeal vocal tract ~100ms after the vocal tract assumed its corresponding configuration. On the other hand, in Spanish, the voiced labial stop [b] normally has an anticipatory phonation onset of ~100ms, meaning that acoustic energy exits the vocal tract ~100ms before the vocal tract assumed its corresponding configuration. Some phonation can be described as almost spontaneous – e.g., the English

voiced labial stop [b], which normally has a delayed phonation onset of only ~20ms. While considerable variation exists in phonation onset duration, this ~20ms threshold cannot be undercut, due to issues of mechanical/aerodynamic delay associated with the morphology of the mammalian vocal tract (Lieberman 1975), and is present in both adult and infant human vocalization as well as that of other mammalian species such as cats (Kiang & Peaker 1960). Thus, “optimal” vocal tract configuration for communication is not possible due to these constraints, but the human vocal tract comes closer to this goal than many other mammalian vocal tracts (Carre 2004; 2009) – it must be kept in mind, however, that the definition of the descriptive term “optimal” is highly subjective in that it can refer to either the greatest possible range of differential phonation or – alternatively – effective communication within one taxon, in which case it would be more relative and taxon-specific. On the flipside, phonation onset can be consciously altered, and thus carry linguistic significance; e.g., in English, the phonetic difference between the labial stops [b] and [p] becomes lost if phonation onset is increased beyond 30ms (Lieberman 1975).

\*\*\*

When talking about linguistic features of human vocalization, it can be said that there is a broad subdivision of human vocalizations into two different categories: vowels and consonants. The distinction between the two is not always clear-cut, and the presence of semi-vowels, also known as glides, makes this readily apparent. Nevertheless, vowels and consonants can be distinguished from each other by some inherent acoustic and phonetic qualities.

Vowels are produced by a two-fold generation of formant frequencies in different parts of the supralaryngeal vocal tract with different cross-sectional area, where F1

originates from the longer part and F2 from the shorter part (Lieberman 1975). Both the horizontal as well as the vertical supralaryngeal vocal tract can assume the role of the longer or shorter part, depending on each vowel's specific configuration. The formant frequencies are differentiated due to air pressure differences at a cross-sectional point  $x$ , where the differentially articulated long and short parts connect to each other – this is usually at the point of intersection between oral and pharyngeal cavity (Lieberman 1975). Because the tongue is one continuous piece of muscular tissue, the range of differential deformation for vowel production is somewhat limited – if the front is pushed in, the back will always be pushed out, and vice versa (Lieberman 1975). While consonants require articulation points for their correct vocalization, most vowels can be produced by anterior-posterior and ventral-dorsal movement of the tongue alone, although some vowels do require some labial manipulation (Lieberman 1975; MacNeilage 2008). Unlike in nonhuman primates, human vowel production is mainly oral instead of nasal, because in humans the velum seals off the nasal cavity during vocalization (Aitchison 2000). To vocalize nasally, it is necessary to actively relax the velum in order to open up the nasal cavity for supralaryngeal air flow (Lieberman 1975).

Extant human languages commonly feature between three and fifteen vowels; out of these, five vowels occur in almost one-third of all languages (29%, Croher 1978; 31%, Maddieson 1984). The most common vowels are [a], [i], and [u] (Trubetzkoy 1939; Liljencrants & Lindblom 1972; Maddieson 1984), which occur in 84%, 92%, and 88% of all human languages, respectively (Maddieson 1984). This is not by coincidence, as these three vowels are also the most phonetically distinct vowels the human vocal tract can produce without losing their distinctive phonemic quality (Lieberman 1975). Each of

these vowels corresponds to an extreme end of possible supralaryngeal vocal tract configuration (Lieberman 1975; Fig. 3). For the vowel [a], the oral cavity is maximally expanded, while the pharyngeal cavity is maximally constricted; the tongue is lowered. For the vowel [i], the oral cavity is maximally constricted, while the pharyngeal cavity is maximally expanded; the tongue is raised. Finally, for the vowel [u], both the oral and the pharyngeal cavity are expanded, while the tongue is raised and the lips are rounded. Additionally, each vowel also corresponds to extremes in formant frequencies. For [a], F1 and F2 are at maximal capacity, while F3 is not. For [i], F1 has a minimal magnitude while F2 and F3 are at maximal capacity. For [u], both F1 and F2 have a minimal magnitude (Boe et al. 2002).

The most critical anatomical feature for the vowel configurations of [a], [i], and [u] is the bent two-tube configuration of the human vocal tract, which inserts three possible constriction points into the vocal tract – one in each cavity, and one at their point of junction – all of which can be acted upon by tongue movement (Lieberman 1975). Due to these maximum configurations, production of the vowels [a], [i], and [u] requires the least amount of articulatory effort, either if vocalized in isolation or within a stream of other phonation (Carre 1996; Lindblom 1998). On the other hand, a single-tube vocal tract as it is found in nonhuman primates and also human infants is not flexible enough for sufficient combined articulation to distinctively produce these vowels (Lieberman et al. 1969).

The vowels [a], [i], and [u], together with labial, dental, and most velar consonants, are called quantal sounds, or phones that are highly distinct from each other, acoustically stable, and easily identifiable by a listener (Lieberman 1991). Again, the bent two-tube



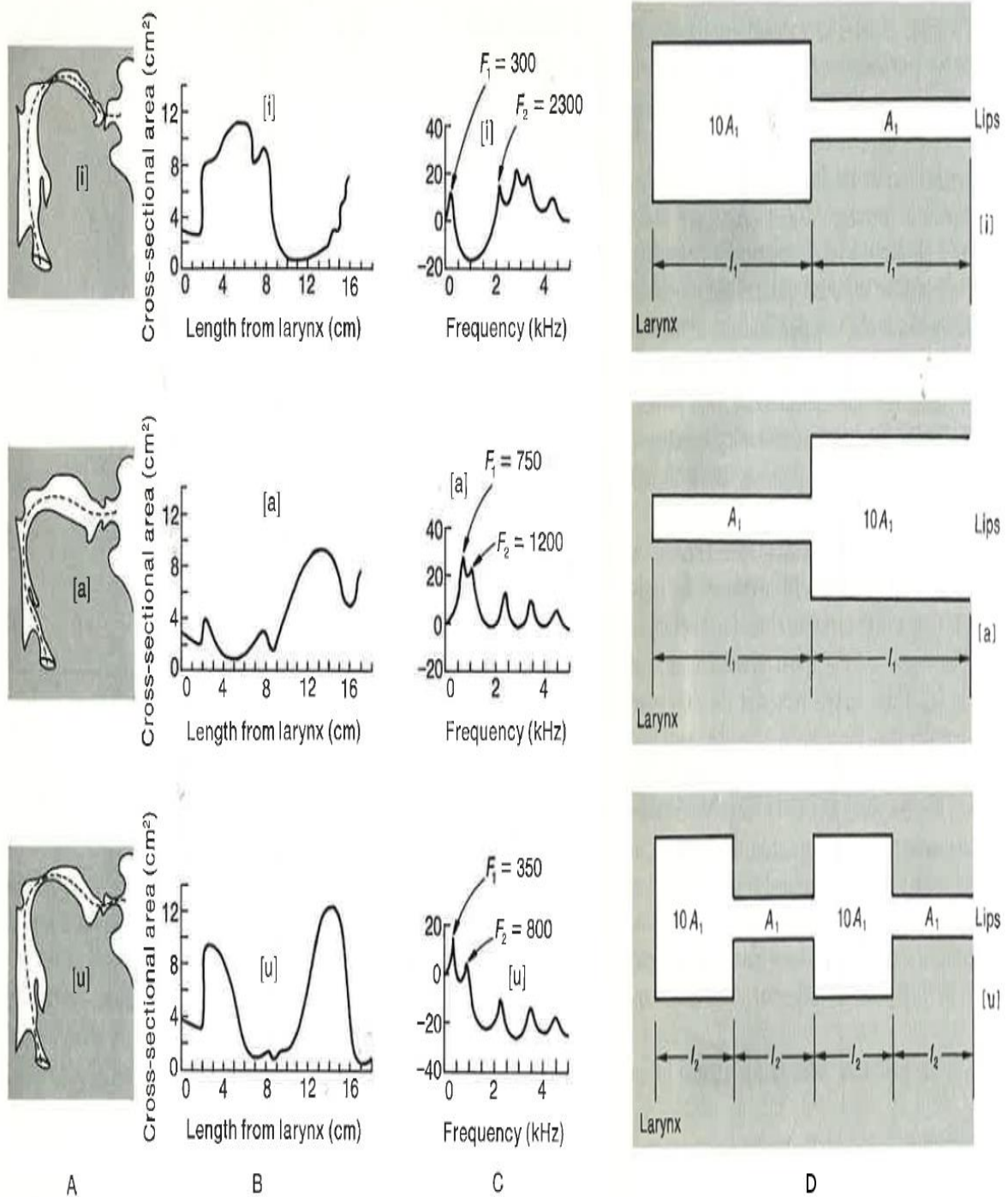


Fig. 3: Cross-Sections of Vocal Tract Configuration (A), Cross-Sectional Area Functions (B), Acoustic Transfer Functions of F1 and F2 (C), and Stylized Two-Tube Vocal Tract Configuration (D) for the Vowels [i],[a], and [u]; modified from Lieberman (1975)

configuration of the vocal tract as well as the morphology of the tongue are critical prerequisites for the successful vocalization of quantal sounds. Quantal sounds occur more often across human languages than any other phones (Greenberg 1963). On the other hand, phones that are easily misidentified by listeners, especially if there is no face-to-face interaction, occur much more rarely if the whole spectrum of human languages is considered (Greenberg 1963). This includes nasalized phonation (Bond 1976), velar stops such as [g] or [k] (Liljencrants & Lindblom 1972), or vowels besides the three quantal vowels [a], [i], and [u] (Peterson & Barney 1952). Interestingly, children tend to learn quantal sounds first, regardless of language background (Jakobson 1968); the order of acquisition of new phones besides quantal sounds then follows the order of commonness of these phones across all human languages, suggesting that quantal sounds' wide distribution could be due to their greater suitability for the vocal tract (i.e., are easier to articulate).

For consonants, the distribution patterns are as follows: stops occur in 100%, nasals and fricatives in 96.8%, liquids in 95.9%, and glides in 90.5% of human languages, based on a survey of 317 different languages, or about 5% of all extant human languages (Maddieson 1984). As such, consonants are much more common than vowels, a fact that is also reflected in the total amount of consonants as opposed to vowels found in these languages, namely 558 and 210, respectively (Maddieson 1984). Out of these 558 consonants, 47% are unique to just one language, and less than 20% occur in more than ten. In the case of consonants, articulation points also play a role in that, say, dental stops might not be universal across all languages even though stops in general are (MacNeilage 2008).

In natural human speech, the smallest unit of sound is not the individual vowel or consonant, but the syllable (Lieberman 1975); in fact, the human vocal tract is unable to produce most phonetic sounds in isolation. For example, one can say [bi], [bu], or [ba], but not [b] by itself. Instead, phonetic sounds are decoded by the brain, allowing for differential perception of [b] from [ba], etc. (Lieberman 1975). While this is true for all consonants, vowels usually can be phonated in isolation, and can thus stand as syllables on their own. Accordingly, the single-vowel syllable (V) constitutes the simplest form of vocalization. If only speech units combining more than one phone are considered, the simplest ones are the consonant-vowel (CV) and the vowel-consonant (VC) syllable, both consisting of a rapid sequence of closed/constricted (consonant) and open (vowel) articulatory configurations (MacNeilage 2008). In almost all these cases, the vowel has a sonority peak, which means that it is louder than any other sound in the syllable (MacNeilage 2008). Lastly, the vowel can be a monophthong or a diphthong, which requires one or two successive tongue configurations, respectively (MacNeilage 2008).

### **Vocal Tract Morphology and Speech Production in Nonhuman Primates**

Next to humans themselves, living primates provide the best clues about the origins of the human vocal tract. This is made possible through a comparative evolutionary systematic approach, which has been successfully employed in the fields of evolutionary biology and paleontology many times. One of the most readily apparent advantages of studying living primates is that all the soft tissue which virtually constitutes the whole vocal tract is present. Additionally, actual vocalization behavior can be observed in live specimen, both in a natural setting in the wild as well as in a controlled setting in

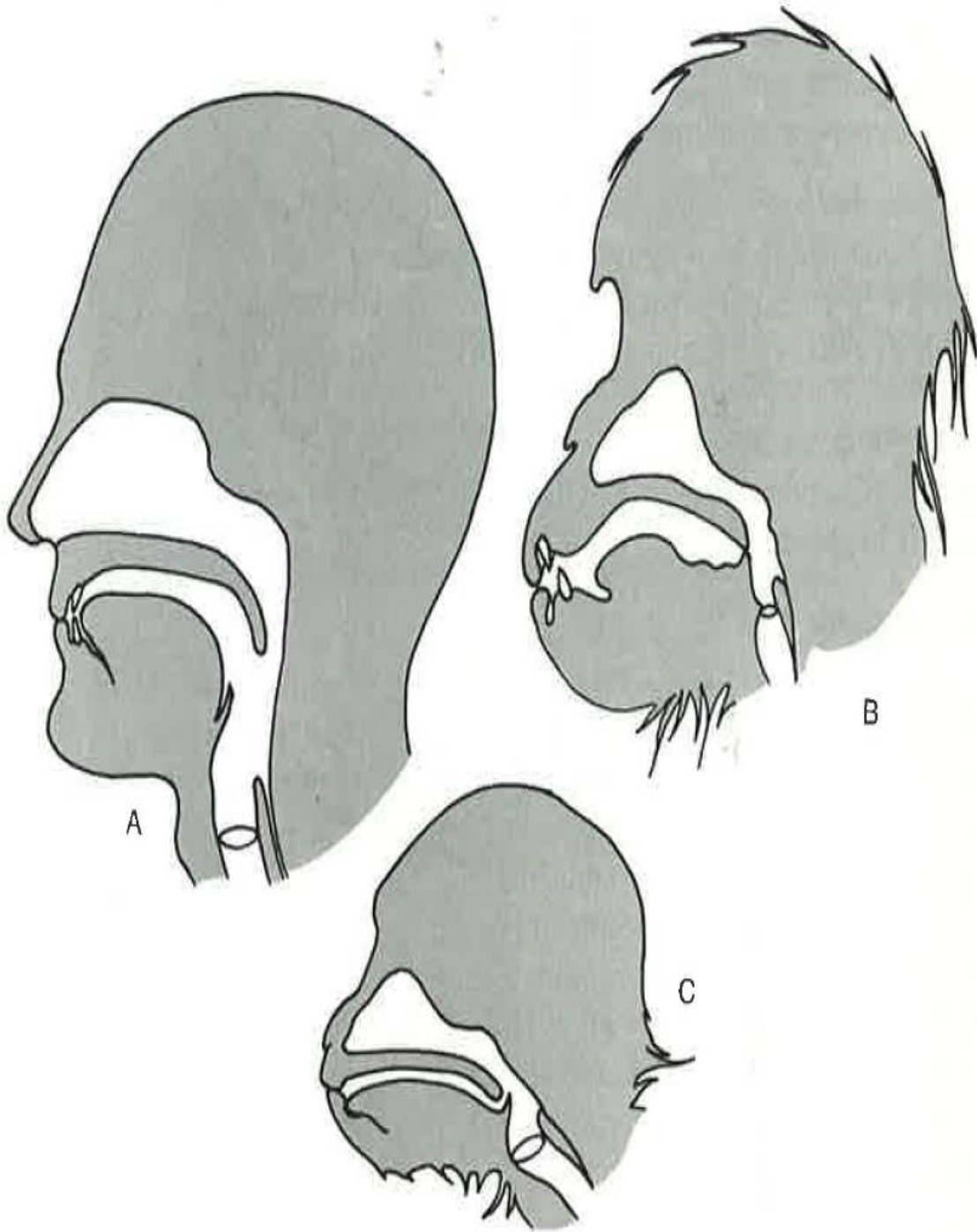
captivity. These advantages were recognized early on in the study of language origins, as dissection of specimens from many different primate species became a favored research objective during the 19<sup>th</sup> and early 20<sup>th</sup> century. Later on, studies of primate behavior also began to be considered in earnest, and more recently, attention has additionally been given to animals outside the primate realm for comparative analysis.

Not surprisingly, the primate species that has been studied most extensively is the common chimpanzee (*Pan troglodytes*) due to its status as humans' most closely-related relative. The logic is that the most common recent ancestor (LCA) of humans with any other primate would have been shared with chimpanzees, and thus morphological similarities between *Homo sapiens* and *Pan troglodytes* should be greater than for any other primate species – with respect to both, as the most closely related ancestors of chimpanzees after bonobos are humans. The reason why bonobos have not received the same degree of attention by studies despite their equal degree of relatedness with regards to humans is that, for quite some time, they were not recognized as a distinct species, and for this as well as other reasons, the amount of data concerning bonobos compared to that concerning common chimpanzees remains relatively low.

One of the most striking differences between human and nonhuman primate vocal tracts is the ratio of the oral cavity to the pharyngeal cavity. In humans, this ratio is ~1:1, whereas in nonhuman primates and other mammals, the oral cavity is larger than the pharyngeal cavity, resulting in a ratio of >1:1 (Negus 1949; Lieberman 1984; Crelin 1987; Flügel & Rohen 1991; Houghton 1993; Laitman & Reidenberg 1993; Nishimura et al. 2003; Nishimura 2005). Additionally, most nonhuman primates show no permanent descent of the larynx and the associated hyoid bone, although there is some slight descent

in a few species such as chimpanzees during maturation (Nishimura et al. 2003; Nishimura 2005), but not to the degree as can be seen in humans, and additionally offset by increased prognathism caused by oral cavity growth (Nishimura 2005). Nonhuman primates' vocal tracts are also not “bent” into a two-tube configuration, as is the case in humans; instead, their vocal tracts are aligned into a continuous single-tube configuration where the oral cavity opens directly into the pharynx, preventing supralaryngeal air flow. As a result, nonhuman primates have a flattened basicranium compared to the complexly-featured human basicranium (Lieberman 1975). Unlike humans, nonhuman primates – and also other mammals – are able to seal off the pharyngeal cavity with their velum due to a heightened position of the epiglottis and the velum, in the process avoiding the risk of choking that humans face (Lieberman 1975). See Fig. 4 for an illustration of these differences.

A major morphological feature that sets humans apart from all other nonhuman primates is the absence of laryngeal air sacs (Fitch 2000b; Ghazanfar & Rendall 2008), although some evidence indicates that vestigial structures might still remain (Ghazanfar & Rendall 2008). The laryngeal air sacs are situated anteriorly to the epiglottis in virtually all nonhuman primates (Negus 1494; Hayama 1970; Hewitt et al. 2002), which pushes the laryngeal structure – larynx and hyoid bone – downward, and also causes the lack of dissociation of the velum and epiglottis that is typical of the human two-tube vocal tract (Flügel & Rohen 1991). The exact number, size, and distribution of laryngeal air sacs vary across different primate species (Ghazanfar & Rendall 2008); in some taxa, such as howler monkeys (*Alouatta*), indris (*Indri*), or gibbons (*Hylobates*), laryngeal air sacs are very well developed (MacLarnon & Hewitt 1999). These taxa also make



*Fig. 4: Two-tube Configuration and Permanently Descended Larynx in Humans (A), and Single-Tube Configuration and Non-Permanently Descended Larynx in Orangutans (Pongo borneus) (B) and Capuchin Monkeys (Cebus) (C); modified from Lieberman (1975), after Negus (1949)*

extensive use of complex vocalization in their natural behavior, and as such it is probable that the laryngeal air sacs play a major role in nonhuman primate vocalization (MacLarnon & Hewitt 1999; Ghazanfar & Rendall 2008).

While many different hypotheses have been made for the purpose of the laryngeal air sacs, such as resonance chambers which amplify long-distance calls, acoustic support for the vocal cords, or extension of the vocal tract for size exaggeration purposes, the most likely explanation for the function of the laryngeal air sacs is that they allow for greater respiratory control by acting as air storage chambers, thus stabilizing vocalization (MacLarnon & Hewitt 1999; Ghazanfar & Rendall 2008). As nonhuman primates lack the advanced respiratory fine-control of humans, they must make use of their laryngeal air sacs to accommodate; however, this mechanism is not nearly as efficient as humanlike respiratory fine-control in terms of vocal endurance. While an average human breath group can last more than 12s (Winkworth et al. 1995), nonhuman primates' breath groups are restricted to 2-6s (Hoit et al. 1994; Mitchell et al. 1996); the longest measured nonhuman primate breath groups on record are those of indris at 5s and howler monkeys at 4.8s. During vocalization, humans extend a normal expiratory phase as much as seven times or more, whereas nonhuman primates usually are restricted to 2-3 times the length; again, indris, howler monkeys, and gibbons come closest to humans with an extended expiratory phase of 4-5 times during vocalization (MacLarnon & Hewitt 1999). While expiration – i.e., the period that can effectively be used for vocalization – makes up 35-65% of the total respiratory phase during vocalization in nonhuman primates, in humans it is as much as 85% (MacLarnon & Hewitt 1999). Lastly, human vocalization sequences are also usually ten times faster than those of nonhuman primates (Lieberman et al 1992),

although all can voluntarily adjust rate of vocalization together with pitch and amplitude if need be (Aitchison 2000).

Laryngeal air sacs tend to be more extensively developed in large-bodied primates (MacLarnon & Hewitt 1999). A possible reason for this is that respiratory phases are proportionate to body size, making fast breathing rates difficult for larger-bodied species (MacLarnon & Hewitt 1999). Accordingly, the maintenance of vocalization at a sufficiently effective rate would become disproportionally energy-intensive; laryngeal air sacs allow large-bodied primates to bypass this physiological limitation (Ghazanfar & Rendall 2008). In humans, who are the third-largest-bodied extant primates after gorillas and male orangutans, this mechanism has been replaced by the even more energy-efficient enhanced thoracic respiratory fine-control. Why this has happened and at what time during human evolution remains unclear; the presence of laryngeal air sacs in both chimpanzees and bonobos (Negus 1949) makes it likely that the LCA of all three species still had these structures. They might have been lost before vocalization became increasingly important for humans, possibly due to behaviorally-induced respiratory morphological requirements that arose when early *Homo* became a long-distance runner (Carrier 1984; Reicholf 1997). Following this assumption, one could speculate whether they would have become important in human vocalization, would eventually have been replaced by thoracic fine-control due to greater efficiency, or actually would have hindered the emergence of modern humanlike speech. This would also mean that, if laryngeal air sacs still had been present in the earlier australopithecines, any kind of hypothetical language or protolanguage in this taxon would have been subject to the same respiratory restrictions as the vocalizations of extant nonhuman primates.



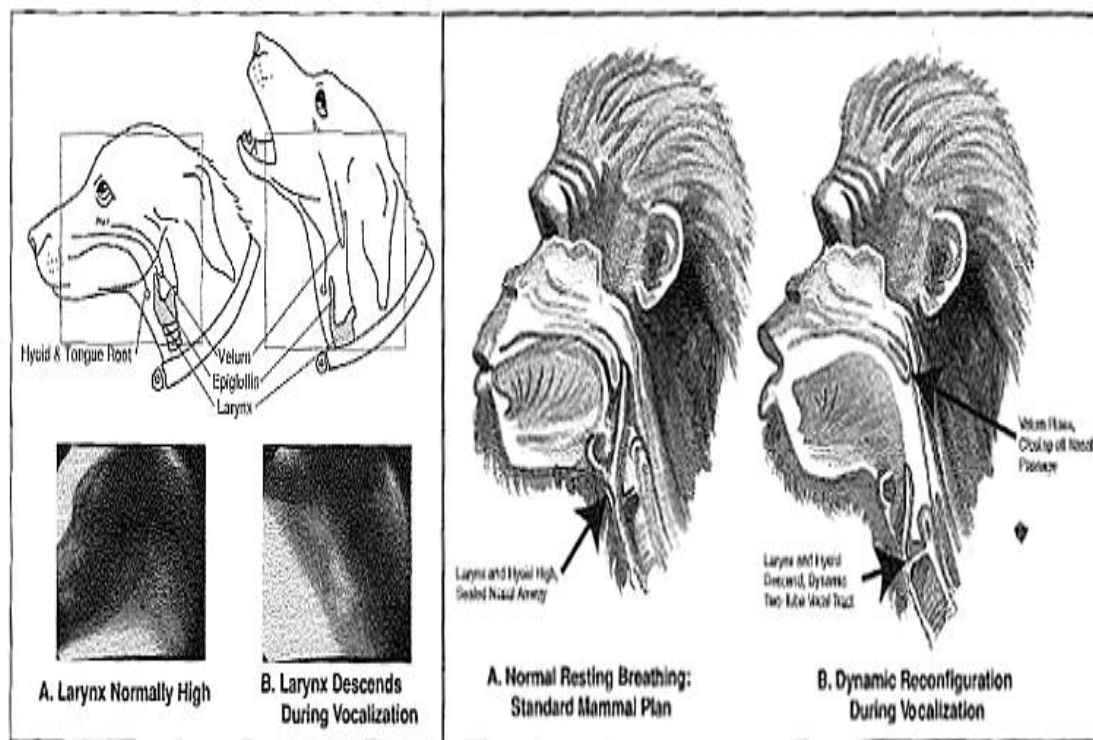
As has been mentioned already, there is no such degree of permanent laryngeal descent in nonhuman primates as there is in humans (Lieberman 1975; Nishimura et al. 2003; Nishimura 2005; Nishimura et al. 2006). For example, in chimpanzees the larynx is situated at a level between the third and the fourth cervical vertebrae, two vertebrae higher than in humans (Lieberman 1975; Barone 1976). However, some slight permanent descent of the larynx also happens in chimpanzees, and the similarities between both processes – including equivalent points of dissociation of the velum/epiglottis and hyoid/larynx (~6 months of age in chimpanzees (Nishimura et al. 2003), ~12-18 months of age in humans (Sasaki et al. 1977)), as well as similar overall pharyngeal growth patterns (114% growth from infancy to adulthood in humans (Fitch & Giedd 1999), and 96% growth in chimpanzees (Nishimura et al. 2006)) – suggest that both laryngeal descent patterns are evolutionarily homologous and would thus have been common in the LCA (Nishimura et al. 2006). Slight laryngeal descent has also been noticed in other primates such as macaques (Flügel & Rohen 1991); however, there is no dissociation between hyoid and larynx in non-hominoid anthropoids (Nishimura 2003), suggesting that macaque laryngeal descent is likely not homologous to that of humans and chimpanzees. On the other hand, the highly integrated anatomical nature of the hyolaryngeal complex makes the presence of such a descent pattern in an anthropoid LCA not entirely implausible, either (Nishimura 2005). In fact, laryngeal descent patterns are more similar in macaques and chimpanzees if compared to humans; however, this is more likely due to human orthognathism –i.e., lack of space for oral cavity expansion – instead of chimpanzee-macaque evolutionary homology (Nishimura et al. 2006).

Descent of the hyoid bone, while virtually identical in human and chimpanzee infants, also varies significantly between adult specimens of each species, with the adult chimpanzee hyoid bone being situated higher than the adult human hyoid bone, even after differences in body size are taken into account (Falk 1975). Instead, chimpanzees' hyoid bone position is similar to that of other apes such as gorillas and orangutans (Negus 1949) as well as gibbons (DuBrul 1958). While the hyoid bone descends during the maturation process in both chimpanzees and humans, it does so less significantly in chimpanzees relative to body size (Nishimura et al. 2003; 2005). There are also differences in timing and rate of hyoid bone descent, which is slower and stops earlier in chimpanzees; maturation of the laryngeal air sacs in chimpanzees, which push the hyoid bone upward during early infancy, is believed to be the cause for this (Nishimura et al. 2006). The human ~1:1 ratio of oral cavity to pharyngeal cavity length is thus mainly due to the greater proportional descent of the hyoid bone compared to chimpanzees and other primates (Negus 1949; Lieberman 1984; Crelin 1987). Additionally, the chimpanzee hyoid bone is relatively larger than the human hyoid bone and has a more hollowed-out corpus. The reason for this feature is that the hyoid bone provides structural skeletal support for the laryngeal air sacs, which extend posteriorly into the hyoid bone in non-human primates (Falk 1975).

The chimpanzee epiglottis, likewise, is situated higher relative to the palate and is wider in diameter compared to the human epiglottis (Jordan 1971). It attaches to the thyroid cartilage (Nishimura et al. 2006), making possible the tight association of epiglottis and velum that gives nonhuman primates the ability to seal off the pharyngeal cavity. For a long time, this morphology was assumed to be the default state, as it was

seen in all autopsies of embalmed specimens (Laitman & Heimbuch 1982; Lieberman 1984; Fitch 2000b; Nishimura 2005); however, more recent magnetic resonance imaging (MRI) studies on live specimens have shown that, at least in chimpanzees, there is a loss of contact in early infancy (Nishimura et al. 2006) similar to the loss that occurs in humans at equivalent stages in the maturation cycle (Sasaki et al. 1977; Nishimura et al. 2003), after which contact must be actively established. Additionally, the epiglottis of chimpanzees and other great apes has an anatomical feature called hiatus intervocalis, which is an ever-present opening with the purpose to relieve high subglottal air pressure during vocalization that would otherwise overwhelm the vocal cords (Kelemen 1948). Humans do not have a hiatus intervocalis; instead, greater articulatory control of the vocal tract musculature provides the vocal cords with the necessary support (Kelemen 1948).

The position of the larynx and its associated structures have long been believed to be universally static (Negus 1949; Lieberman 1984; Crelin 1987; Flügel & Rohen 1991; Houghton 1993; Laitman & Reidenberg 1993), with accordingly any vocal implications stemming from larynx height being fixed. However, X-ray imaging studies in the last two decades have revealed that this is not the case; in fact, many mammals – primates and non-primates alike – are able to retract their vocal apparatus temporarily via the sternothyroid strap muscles from its higher resting position while simultaneously raising their velum, and thus can achieve a humanlike laryngeal configuration during vocalization (Fitch 2000b) (Fig. 5). This ability already has been observed in pigs, dogs, goats, and some New World monkeys such as tamarins (Fitch 2000b). Consequently, any studies on vocal anatomy based on dead specimens – as is the case during an autopsy –



*Fig. 5: Temporary Descent of the Larynx during Vocalization in Dogs and Primates; modified from Fitch (2010)*

likely cannot convey the full theoretical vocal configuration potential of many mammals (Fitch 2010). This fact especially must be kept in mind when turning to the fossil record, underlining the importance of the comparative approach specifically in a case such as this one, where there is no readily discernible association between soft tissue and skeletal features (Fitch 2002), meaning that no traces that can be used for fossil reconstructions are present.

At the same time, even though these mammal taxa are able to accommodate laryngeal position for rapid formant frequency shifts like humans do, all of them can vocalize without doing so, and none is documented to actually make use of this capability in their

natural vocalization (Fitch 2000b). The primary purpose of laryngeal descent in these species is probably more in order to increase vocalization volume – i.e., sound louder – than any other reason, and as a matter of fact, the reason why they do not produce humanlike formant frequencies could well be caused by neural signaling deficiencies instead. As such, it seems that only in humans has the phonatory subsystem of the vocal tract specifically and purposefully evolved to make use of this potential vocalization possibility for articulatory as well as acoustic stabilization purposes. Support for this assumption comes from the numerous differences between humans with a static larynx and other mammals that can temporarily lower their larynges, such as an inability vs. an ability to connect larynx and velum to prevent choking, a thick and rounded vs. a thin and flat tongue, and a bent two-tube vs. a single-tube vocal tract, respectively (de Boer 2010). Because chimpanzees also can retract their larynx slightly during vocalization (Nishimura 2005), it can be suggested that the uniquely human innovation within the primate lineage is a static permanently-descended larynx.

Another recent breakthrough in the field of comparative vocal tract anatomical studies has been the revelation that humans are the only primate, but not the only mammal species with a permanently descended larynx. As a matter of fact, koalas had been known to have a permanently descended larynx very similar to that of humans as early as the 1920s (Sonntag 1921); unfortunately, Sonntag's study never received much attention, and thus his findings were remained relatively unknown until fairly recently (Fitch 2010). More recently, however, a permanently descended larynx has been attested in other mammal species as well, including Mongolian gazelles, where it only occurs in males (Frey & Riede 2003); red and fallow deer, where it also only occurs in males and is not

associated with the hyoid bone, making it possible to temporarily retract the larynx even further all the way to the bottom of the pharyngeal cavity during vocalization (Fitch & Reby 2001); Arabian camels, similar to deer (Fitch 2010); and big cats of the genus *Panthera*, which like humans have a tightly associated larynx and hyoid bone attached to the skull only via an elastic ligament (Weissengruber et al. 2002). The relatively simple vocalization patterns of all these species indicate that laryngeal descent's primary evolutionary advantage is not vocalization (Fitch 2002).

A possible answer to this puzzling phenomenon is found in the size exaggeration hypothesis (Ohala 1984; Fitch 2002). The size exaggeration hypothesis states that mammals use fundamental as well as formant frequencies to correctly estimate the speaker's body size via deduction of vocal tract length. This works especially well with quantal vowels and specifically [i] (Lieberman 1975; Fitch 1994), which for this reason is also called a super-vowel (Nearey 1978). Other phones, such as semivowels and glide consonants, are also suitable for size deduction (Lieberman 1975). Size deduction is critical for vocalization purposes, as each individual vocal tract produces different frequencies due to differences in vocal tract size (Lieberman 1975) – the larger the vocal tract, the lower the frequencies – with different phonation for the same intended phones as a result, and sometimes even allowing for different articulatory means to do so (Lieberman 1975). As such, the listener must “calibrate” the speaker's vocal tract against a theoretical norm, which is done by estimating vocal tract length. This calibration mechanism is based on frequency and occurs in many mammal species, both primates (Fitch 1997) and non-primates (Riede & Fitch 1999).

Size can be exaggerated, however, through distortion of these frequencies via a lowered larynx (Fitch & Reby 2001). The greater the laryngeal descent in an individual vocal tract, the lower the formant frequencies it produces will be. Being able to manipulate these formant frequencies and thus making an appearance of larger-than-actual body size would thus prove evolutionarily advantageous, as it can be used to impress potential mates and also intimidate rivals (Lieberman 1975; Krebs & Dawkins 1984). A permanently descended larynx fits this purpose in the aforementioned species of deer, camels, and gazelles, where it occurs only in males; however, its role is slightly more complicated in humans, because it appears in both sexes and has an undeniable importance for human speech production. Interestingly though, in human males, a second phase of laryngeal descent occurs during puberty, which does not happen in females (Ohala 1984; Fitch & Giedd 1999; Fitch 2002; Vorperian et al. 2009). In fact, laryngeal descent is one of the most sexually dimorphic features that exists in *Homo sapiens*; while males are on average only 10% taller and 20% heavier than females, the average adult male's F0 is only half that of the average adult female's F0 (Ghazanfar & Rendall 2008). Size estimation in humans is mostly based on F0 instead of F1, F2, and F3, even though F0 is less reliable for such purposes (Ghazanfar & Rendall 2008) – which is most likely on purpose, since the goal is to deceive others into overestimating one's body size. Correspondingly, humans instinctively assume that there is a correlation between body size and voice pitch (Smith et al. 2005) – determined by F0 – even though no clear-cut correlation such as this exists in adult humans (van Dommelen 1993). Even more intriguingly, in terms of behavioral adjustment human males tend to value voice-deduced body size more than human females, a fact that strongly suggests that size exaggeration

evolved predominantly as an element of male-male dominance rivalry instead of male-female mate selection (Puts et al 2006). This is not to say that human females do not value voice-deduced body size at all; if there had been no preferential mate choice by females for males with a lower F0, no intersexual selective pressure on the prevalence of this trait would have occurred. At the same time, however, a dominant male would have been able to prevent other males from mating, thus contributing to the intrasexual selection of this trait. Because a dominant male would likely have had greater reproductive success for both reasons, there would have been sexual selective pressure for additional laryngeal descent in adult males.

Because the descended larynx is important in human speech production, and because initial laryngeal descent also occurs in human females, the role of the descended larynx probably expanded from size exaggeration purposes to communicative competence at some point during human evolution (Fitch 2002). As was mentioned in section one, the adult human male larynx is at a less optimized position for articulation than the adult human female larynx. According to the size exaggeration hypothesis, the reason for this is that in human males, laryngeal descent was stretched to the maximum capacity possible for size exaggeration purposes, while still maintaining sufficient articulatory ability; adult human females, which are free from this evolutionary selective stress, evolved to have an optimally-situated larynx instead (de Boer 2010). Differential sexual selective pressure is thus likely responsible for the over-proportional degree of sexual dimorphism of the human larynx.

Moving the focus of the discussion back to nonhuman primates, the most apparent feature that distinguishes their supralaryngeal vocal tract is, again, the relatively greater



size of the oral cavity compared to the pharyngeal cavity, which causes the aforementioned >1:1 ratio of nonhuman primates compared to the ~1:1 ratio in humans. However, these ratio differences are not present at birth, but gradually emerge during the maturation cycle (Nishimura et al. 2006). At birth, both human and chimpanzee infants have ratios of ~1.5:1, meaning that the oral cavity is also much larger in human infants in comparison to the pharyngeal cavity. While this imbalance is offset by rapid growth of the pharyngeal cavity and less growth of the oral cavity (oral cavity length in human infants: ~5cm at birth, ~7cm at nine years of age; pharyngeal cavity length: ~3.5cm at birth, ~7.5cm at nine years of age), a similar growth pattern of the pharyngeal cavity also occurs in chimpanzee infants; however, the oral cavity grows much larger than it does in humans (oral cavity length in chimpanzee infants: ~5.5cm at birth, ~9-10cm at five years of age; pharyngeal cavity length: ~2.5-3cm at birth, ~4.5cm at five years of age) (Nishimura et al. 2006). The >1:1 supralaryngeal vocal tract ratio in chimpanzees is thus mainly established by greater growth of the oral cavity.

A possible cause for this greater size of the oral cavity relative to humans is that the tongue, which is the most important articulatory feature of the supralaryngeal vocal tract, is situated entirely in the oral cavity of nonhuman primates, whereas in humans it also forms the anterior wall of the pharyngeal cavity (Lieberman 1975; Lieberman 1984; Crelin 1987; Zemlin 1988; Titze 1994). This morphology effectively removes two of the three constriction points that characterize the human two-tube supralaryngeal vocal tract and thus restricts potential cross-sectional area modifications necessary for rapid sequential vocalization (Lieberman et al. 1969; Lieberman 1984; Fitch 2000b; Fitch & Hauser 2002). In turn, this greatly reduces nonhuman primates' vowel range – at least

with regard to the quantal vowels [a] and [i] (Lieberman 1969; Lieberman 1975) (Fig. 6). Because the quantal vowel [u] does not require maximum expansion of either the oral or pharyngeal cavity, a single-tube supralaryngeal vocal tract is theoretically sufficient for humanlike phonation of [u], but the presence of the simian shelf – an internally buttressing bony protrusion of the mandible – in nonhuman primates prevents the required articulatory movement of the tongue in the oral cavity (Lieberman 1975).

The small body size of many primates is another obstacle for the successful phonation of quantal vowels, which require sufficiently large cavities to generate enough turbulent air flow for the necessary formation of quantal vowel formant frequencies (Lieberman 1975). Even large-bodied primates such as gorillas, orangutans, chimpanzees, or baboons lack a humanlike quantal vowel space due to the limitations on formant frequency stabilization that are inherent to the single-tube vocal tract as well as the simian shelf (Lieberman 1975; Owren et al. 1997). Another factor might be the lack of articulatory fine-control required to coordinate the supralaryngeal vocal tract and larynx in nonhuman primates, which has been linked to the FOXP2 gene; this gene, at least in its human form, is not present in other primates (Zuberbühler 2005).

Despite all these differences in vocal tract anatomy, human and nonhuman primate vocalization shows some remarkable similarities both in acoustic as well as articulatory terms. Philip Lieberman (1975:97-100) has compiled an extensive list of these similarities. According to him, nonhuman primates can, and often do:

- produce voiced and unvoiced phonation
- phonate at high frequency, normal frequency, and low frequency

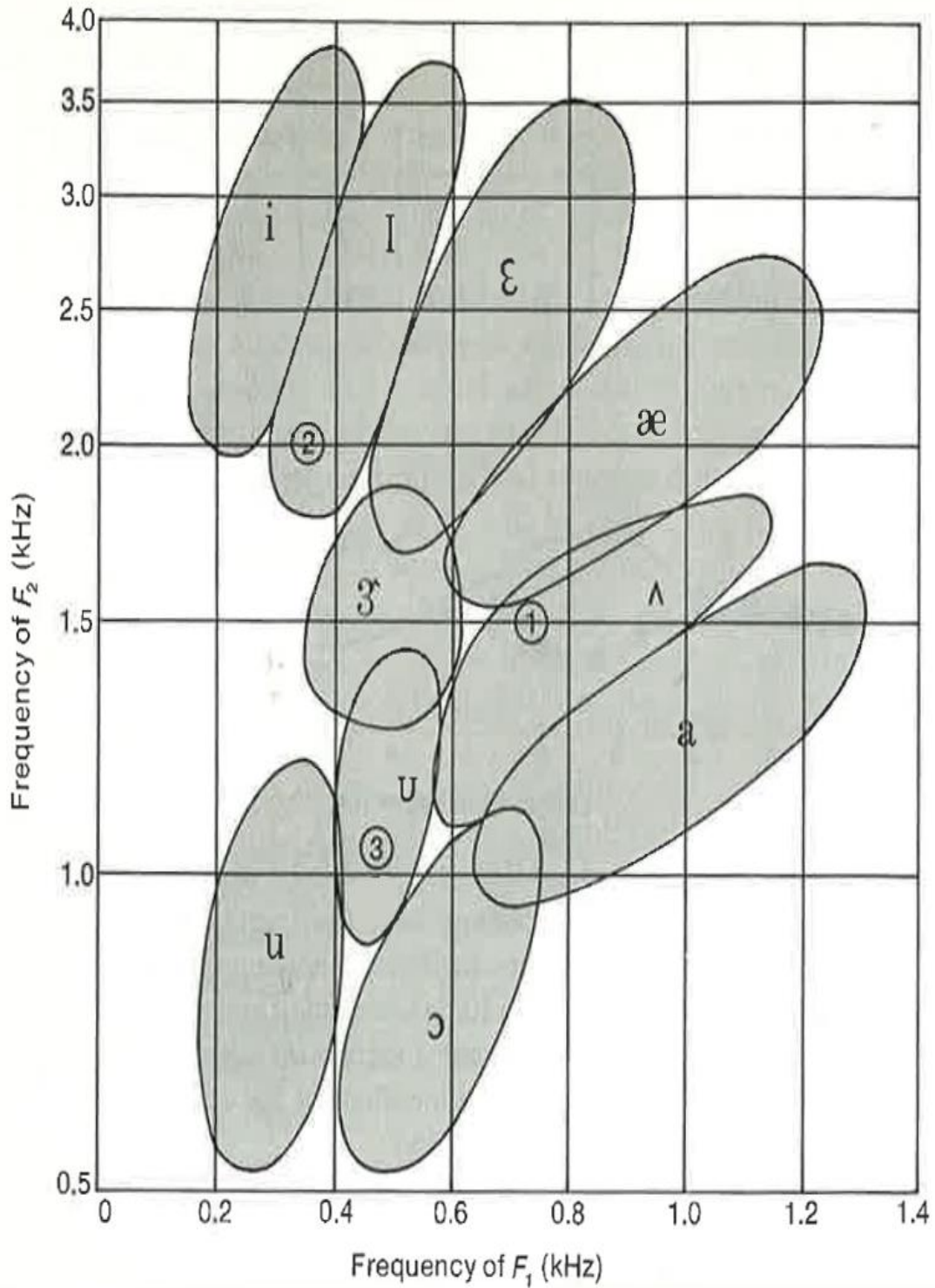


Fig. 6: Human Vowel Space in terms of F1 and F2, and Closest Chimpanzee Capability to Phonate [a] (1), [i] (2), and [u] (3); modified from Lieberman (1975)

- adjust their formal frequency variation, and thus are able to produce paralanguage (pitch, tone, “breathy” language)
- make extensive use of breath groups
- filter air flow using their supralaryngeal vocal tracts
- produce stops
- potentially use bilabial, dental, and glottal articulation points (it is not actually clear if they have the neural/muscular fine-control to extensively make use of these; their most reliably produced consonants are dental, and these are also the first consonants mastered by human infants (Aitchison 2000); non-human primates also lack alveolar and velar articulation points)
- transition between formant frequencies during ongoing vocalization, mainly by articulation of the lips instead of the larynx, which is the case in humans
- produce non-oral nasalized vocalization (humans can only orally produce nasalized vocalization)

In great likelihood this list is incomplete, as nonhuman primates certainly do not make use of their full potential vocal ability in their natural vocalization behavior, similar to how no single human language encompasses the entire phonation range humans can produce in its native phonemic repertoire (Lieberman 1969). As an example, Japanese macaques are fully capable of employing the labial articulation point, but almost never do so in their natural vocalization, instead relying on other articulation points (Itani 1963).

On the other side, there are also a number of things nonhuman primate vocalization definitely lacks in comparison to that of humans. Nonhuman primates cannot:

- maintain stable formant frequencies over the extent of a breath group without large fluctuation, which would make any hypothetical speech on their part sound incomprehensible, or at least coarse and unpleasant to human listeners (Lieberman 1975)
- achieve the same vowel space as humans, due to limitations of the single-tube vocal tract in terms of maximum expansion (but not constriction) (Fant 1960; Lieberman 1969)
- produce stable quantal vowels (Lieberman 1975)
- produce non-nasalized vocalization, due to anti-resonances – frequencies at such magnitudes that acoustic energy is effectively absorbed – which are caused by their pharyngeal morphology (Lieberman 1975)
- laugh, because of an inability to separate a single expiratory breath phase into separate sound units (Provine 1996); however, they could produce laugh-like vocalization during inspiration
- produce humanlike formant frequencies (Fitch 2002), although they are latently able to perceive them (Fitch & Kelley 2000)

In other cases, non-human primates are theoretically able to produce some humanlike phones but lack the necessary neural prerequisites to do so (Lieberman 1991).

Broadly speaking, nonhuman primates make relatively less use of their articulatory subsystem than humans in their natural vocalization (MacNeilage 2008). This is not to say that nonhuman primates do not produce very complex vocalizations themselves; indeed, many species of Old World Monkeys have developed elaborate call systems featuring calls that are phonetically very distinct (Zuberbühler 2005), sometimes

complementing these call systems with additional gesturing – e.g., Rhesus macaques (Reynolds 1972). Similarly, gibbons have highly complex, multi-syllabic calls with specific syllabic arrangements that require meticulous articulatory manipulation (Brockelman & Schilling 1984; Geissmann 1984). Still, just like humans most non-human primates do not fully exploit the entire repertoire of vocalization available to them, and computer-modeling studies indicate that even if they did, their vocalization range would fall short of that which is used by humans in normal language – let alone the entire human range of vocalization across all languages (Lieberman 1969; de Boer 2010).

Gibbon vocalization raises another issue of importance in nonhuman primate vocalization: like that of other nonhuman primates, it is an instinctual behavior. Gibbon calls are population-specific, and hybrids from separate populations produce corresponding hybrid calls (Geissmann 1984). Likewise, gibbons raised in captivity still vocalize like their contemporaries in the wild (Brockelman & Schilling 1984). Generally speaking, nonhuman primates are entirely capable of vocalizing if they wish to, but they seemingly are unable to *not* vocalize when they are presented with a stimulus which would induce such vocalization (Lieberman 1991; Aitchison 2000). Jane Goodall has reported that chimpanzees at Gombe, when presented with bananas, instinctively vocalize, which draws other chimpanzees to the site, putting the calling chimpanzee at a disadvantage because he or she now has to share the food with the other members of his or her group (although this behavior could also be explained as altruism, which would be advantageous of the altruistic individual because he or she can expect some favor in return). When a young male who otherwise was often excluded from the feast due to his low rank within the group was given some bananas, he had a very hard time trying to

suppress this sort of vocalization; eventually, he failed, and again had to share most of his bananas with the higher-ranking other chimpanzees (Goodall 1986). As such, one definitive feature setting apart human language from nonhuman primate vocalization is this inability to dissociate vocalization from its corresponding stimulus, and consequently the ability to use that vocalization in an abstract setting with or without personal gain (Aitchison 2000).

One explanation that has been offered to explain this phenomenon is that human speech is tri-phasic, whereas most other mammals rely on bi-phasic vocalization (MacNeilage 2008). Bi-phasic vocalization is essentially a cooption of the chewing and/or suckling reflex – i.e., it resembles a steady, rhythmic sequence of opened-closed mouth configurations. Humans, on the other hand, vocalize in unpredictable open-closed sequences, which is tri-phasic. Tri-phasic vocalization is almost certainly a simple modification of bi-phasic sequencing, according to the principle of parsimony – less evolutionary effort was required to modify an existing mechanism rather than to innovate an entirely new one. At the same time, some tri-phasic mimic gesturing is actually present in nonhuman primates in the form of lip smacks, which are regularly performed in social settings (Redican 1975) and often accompanied by some sort of vocalization (Green 1975; Andrew 1976). It is thus possible that lip smacks and similar forms of mimic gesturing could have acted as a sort of behavioral precursor to humanlike speech (MacNeilage 2008).

\*\*\*

Many attempts have been made throughout the history of science to go beyond nonhuman primates' natural vocalization by teaching them human speech-based

language. As early as 1747, the French philosopher Julien Offray de La Mettrie, drawing on René Descartes “bête machine”, argued that chimpanzees could be taught to speak like humans if one accounts for their neural condition – i.e., teaching them the way one would teach a mentally retarded, language-impaired child (Lieberman 1975). Many subsequent attempts were made using this unflattering methodology, but none was successful (Kellogg 1986) – a trend that has held true to this day.

Beginning in 1951, the scientist couple Keith and Catherine Hayes began rearing a chimpanzee infant named Viki, together with their own infant son (Hayes & Hayes 1951). They hoped that the same socialization context would enable Viki to speak like a human child; however, while the Hayes’ son picked up his parents’ language normally like any human child *and* additionally learned to vocalize in a similar way to a chimpanzee from his simian playmate, Viki never managed to say more than four distinct English words (“mama”, “papa”, “up”, “cup”), even though the Hayes’ actively tried to help her do so by manually manipulating her vocal tract.

Learning from this experience, subsequent researchers instead focused on teaching their ape subjects sign languages. This fared much more successfully, as chimpanzees like Washoe (Gardner & Gardner 1992), Sarah (Premack & Premack 1983), Nim Chimpsky (Terrace 1979), and Austin & Sherman (Savage-Rumbaugh & Lewin 1994); gorillas like Koko (Patterson 2001); orangutans like Aazk (Shapiro 1982); or bonobos like Kanzi (Savage-Rumbaugh & Lewin 1994) all learned to use dozens or even hundreds of signs, and sometimes also mastered rudimentary grammar (all sources cited in Desalles 2000). None of these attempts was without their difficulties, as often even very simple signs had to be repeated hundreds of times to the apes – combined with the



necessity of rewarding – before they successfully learned them (Desalles 2000). In some cases, serious setbacks occurred; e.g., in Nim Chimsky's case, it eventually turned out that he might have just mimicked his teachers, but failed to actually understand what he was doing. Other cases went beyond all expectations; the bonobo Kanzi, in particular, seems to be a natural prodigy, becoming probably the most prolific nonhuman great ape communicator ever, knowing more than 1000 signs. What was most impressive was that Kanzi mostly taught these signs to himself without any human input, and also learned to understand some simple spoken English. It is certainly possible that Kanzi does not reflect the language capabilities of an average bonobo, and instead could be an exceptionally gifted individual. This becomes especially apparent when considering that his mother, the actual subject of the study who was taught sign language for two years, learned only six signs after more than 30,000 attempts (Desalles 2000). Yet even Kanzi never learned to speak English, or any other human language (Desalles 2000; Fitch 2010).

While this might lead one to suspect that great apes do not speak because their vocal apparatus is insufficiently equipped for such an endeavor, it is not the whole story, and to find it, one must look outside the primate and mammal realm. The animal studies on language that have rocked the scientific world to an almost comparable degree as the great ape studies were those undertaken by Irene Pepperberg on African grey parrots. The two parrots Pepperberg studied, Alex and Griffin, were capable of learning many different English phonemes and combining them independently into new vocal constructs (Pepperberg 2005) – e.g., Alex used “grey” to independently form “grain”, “grate”, and “grape”. Both Alex and Griffin's vocalization featured allophones – e.g., the /k/ in “key”

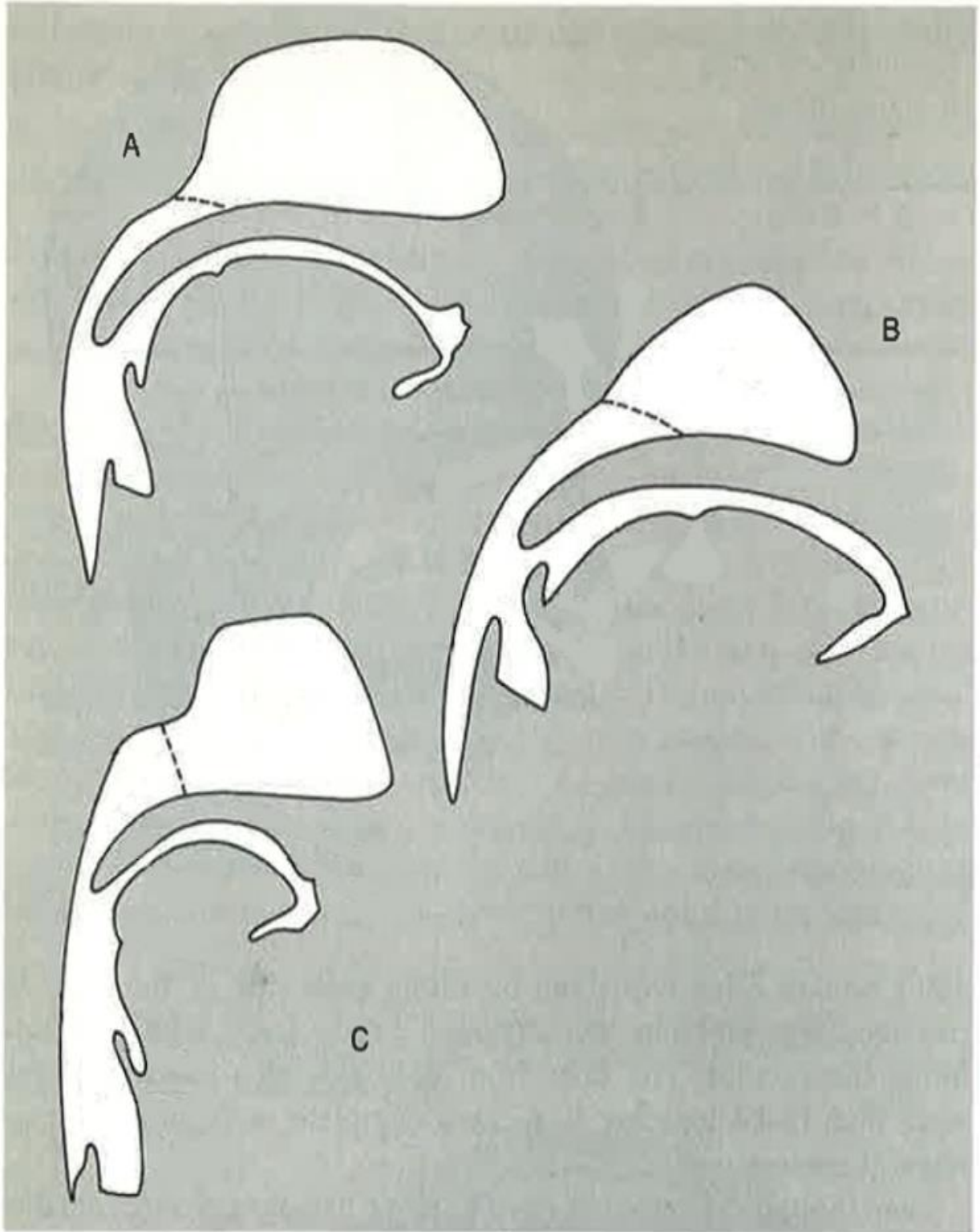
sounded distinctly different than the /k/ in “cork” (Patterson & Pepperberg 1998). Alex also was able to recognize and produce minimal pairs, such as /t/ea and /p/ea, or cor/k/ and cor/n/ (Pepperberg 1999). While the parrots were not able to learn nearly as many different meanings as great apes, they did understand their meaning and were able to use them in linguistic contexts accordingly – and, unlike in the case of great apes, there was little disparity between what could be understood and what could be verbally communicated (Pepperberg 2005).

Due to the fundamental differences in mammalian and avian vocal tract morphology, naturally parrot vocalization differs substantially from those of primates – e.g., Alex and Griffin equally used their tongues and beaks (roughly equivalent to the lips) for articulatory configuration, whereas primates – including humans – primarily use the tongue instead of the lips. In other instances, however, there are similarities – e.g., parrots also have phonation onset and anticipatory coarticulation in that they anticipate upcoming phonation and adjust their vocal tract accordingly during vocalization (Warren et al. 1996). The success of these parrots in both actively understanding and employing as well as actually vocally articulating human speech – an area where great apes have failed (Wray 2005) – raises the question whether humanlike vocalization actually requires a humanlike vocal apparatus (Pepperberg 1999), and whether the language threshold in nonhuman primates is thus more likely neurologically determined – e.g., an inability to control the articulatory musculature in ways required for successful humanlike vocalization due to neural signaling deficiencies for such purposes (Fitch 2010).

## **Vocal Tract Morphology and Speech Production in Human Infants**

Besides the study of extant nonhuman primates, another promising scientific approach that could lead to a greater understanding of human vocal tract morphology and speech production capabilities is the study of human newborns' and infants' vocal tracts. The rationale behind this is based in fundamentals of embryology stating that, morphologically speaking, all new vertebrate life is virtually the same at conception (Shubin 2008). Morphological taxon-distinctive features subsequently arise during embryological and later fetal development in the womb; in many cases, this differentiation continues after birth in infants and juveniles, only reaching full completion in adult specimens – i.e., the maturation process. This especially holds true for humans, who are born at a relatively premature stage compared to other mammals because the large human cranium would become too big for the mother's birth canal at more developed stages (Shubin 2008). Consequently, alongside many other features, a human infant's vocal tract also differs substantially from a fully mature one. The maturation process of the vocal tract is generally complete by nine to fifteen months of age, when human children acquire the full range of human vowel vocalization (Winitz 1960).

Generally speaking, a human infant's vocal tract resembles a nonhuman primate vocal tract more than an adult human vocal tract, in that its configuration is a single-tube instead of a bent two-tube (Lieberman 1975). Accordingly, a number of morphological features associated with such a vocal tract configuration which are present in nonhuman primates also can be found in human infants (Fig. 7). Most prominently, the human infant oral cavity to pharyngeal cavity ratio is  $>1:1$ , with the oral cavity being significantly larger than the pharyngeal cavity, which is caused by the lack of laryngeal descent in



*Fig. 7: Human Infant (A) and Chimpanzee (B) Single-Tube Vocal Tract, and Human Adult (C) Bent Two-Tube Vocal Tract; modified from Lieberman (1975)*

newborns (Negus 1949; Lieberman 1984; Crelin 1987; Lieberman et al. 2001). As was discussed before, pharyngeal cavity growth is much more pronounced than oral cavity growth during maturation, with pharyngeal cavity growth being twice as great as oral cavity growth (Goldstein 1980). The main cause of this is laryngeal descent, which begins around three months of age and is finalized around four years of age (Vorperian et al. 2009); the adult human ~1:1 ratio of oral cavity to pharyngeal cavity is established in the process (Negus 1949; Crelin 1987; Fitch & Giedd 1999; Vorperian et al. 2005). This developmental change allows the pharyngeal cavity to function as a resonator of equivalent magnitude to the oral cavity (Lieberman 1984; Crelin 1987; Titze 1994; Fitch 2000), and thus makes possible the formation of the necessary formant frequencies for humanlike speech. Growth rates of oral cavity and pharyngeal cavity also vary in duration, as the oral cavity reaches mature size long before the pharyngeal cavity (Vorperian et al 2009).

The anatomical processes causing laryngeal descent in human infants include a descent of the hyoid bone relative to the static hard palate, as well as additional descent of the laryngeal skeleton relative to the hyoid bone (Fitch & Gied 1999; Lieberman & McCarthy 1999; Lieberman et al. 2001; Vorperian et al. 2005). Interestingly, throughout this elongation process of the pharyngeal cavity and the supralaryngeal vocal tract in general, formant frequencies remain relatively stable, without the acoustical instability associated with nonhuman primate vocalization (Kent & Murray 1982; Robb et al. 1997; Gilbert et al. 1997). The large increase in pharyngeal volume and consequently unchanged resonance could explain this occurrence (Vorperian et al. 2009). However, because formant frequency magnitude is inversely proportional to vocal tract length, in

human infants – whose vocal tract length is ~7cm compared to ~15cm and ~18cm in adult human females and males, respectively (Vorperian et al. 2009) – the formant frequencies tend to be much higher than in adults. Consequently, human infants' range for formant frequency variation is very limited in comparison to that of adult humans (Truby et al. 1965), but similar to that of non-human primates such as chimpanzees or gorillas (Lieberman 1975). But unlike human primates, during language acquisition human infants innately take into account the smaller size of their vocal tracts compared to those of adults by assuming formant frequencies with similar ratios to those employed by adults; the human ability to estimate vocal tract size based on formant frequency ratio then allows adult humans to “calibrate” infant speech in order to correctly perceive it (Lieberman 1991). Starting around two years of age, formant frequencies begin to fall as would be expected relative to vocal tract length, and in fact slowly continue to do so throughout life for both genders, suggesting that there is some slight continuous morphological change in pharyngeal cavity dimensions even after maturity (Vorperian et al. 2009).

As was noted in section two, laryngeal descent and pharyngeal cavity growth are also present in chimpanzee infants and possibly other nonhuman primates, but not to the extent as happens in human infants. Drawing on these similarities, it seems likely that such maturation growth trends evolved gradually as demands for increased vocalization complexity and frequency increased in primates (Fitch 2000b; Fitch 2002). However, pharyngeal cavity elongation and two-tube supralaryngeal vocal tract configuration were probably independently-arising features (Lieberman 1975; Stevens 1989), with the latter arising after the human-chimpanzee split due to its absence in the maturation cycle of

chimpanzee infants (Nishimura et al. 2006). Because the laryngeal complex and the pharyngeal cavity originate from different tissues of diverse embryological origins – as is generally the case for the whole vocal tract and other parts of the head (Sadler 2006; Shubin 2008) – which corresponds with different growth onset phasing in the maturation cycle (Vorperian et al. 2009), such a differential evolutionary origin seems even more likely. Accordingly, the retention of a nonhuman primate-like vocal tract in newborn human infants can be understood within this context. There are also significant evolutionary incentives associated with such developmental delay; because newborn infants still lack the neural capacity to make use of a fully functional vocal tract by means of language, postponement of its development enables infants to avoid the risks to survival associated with an adult human vocal tract configuration, such as choking (Lieberman 1991).

A number of other morphological differences between human infant and adult vocal tract configuration exist. In all these cases, the human infant vocal tract more closely resembles that of non-human primates than that of its adult counterpart. Again, Philip Lieberman (1975:108-113) has compiled a list of these features:

- the ability to seal off the oral cavity with the epiglottis and thus prevent air flow, which is impossible for adult humans
- a direct continuous opening of the oral cavity into the pharyngeal cavity, without an intervening laryngeal space
- connection of the hyoid and cricoid cartilages, restricting laryngeal movement during vocalization to ~5mm compared to ~20mm in adult humans, which exceeds the difference expected even if body size differences are considered

- no inferior extension of the anterior portion of the geniohyoid and digastric muscles, due to a higher positioning of the hyoid bone
- inward instead of outward flaring of the mandible, as the associated muscle attachment sites are brought into correspondence with the above-mentioned muscles
- less vertical inclination of the styloid processes, which serve as muscle attachment sites for the stylohyoid muscles, which in turn support the hyoid bone
- a pushed-back sphenoid bone in order to make room for the higher-positioned larynx; this – coupled with human orthognathism and bipedalism, the latter of which has major impacts on cranial and basicranial layout – in turn requires extensive cranial restructuring during maturation, a process that is distinct from those of all other mammals (Laitman & Crelin 1976)
- a flattened basicranium, caused by a single-tube vocal tract configuration

These morphological differences between the human infant and adult vocal tract influence human infant vocalization substantially, even though, of course, it is not yet speech-based – i.e., human infants do not actively and consciously use vocalization to convey meaning – and vocal tract maturation reconfiguration is usually complete by the time children extensively begin to use language. On the other hand, the great similarities between nonhuman-primate and human infant vocal tracts are responsible for the overall similarities between their vocalizations, although there are some differences – e.g., like nonhuman primates, human infants lack the required articulatory fine-control for the application of sufficient muscular tension to the larynx during vocalization in order to achieve humanlike speech (Lieberman 1975). However, because human infants have no



hiatus intervocalis, their vocal cords collapse under the air pressure flowing from the open glottis (Lieberman 1975), causing the “noisiness” associated with human infant cries. In another example, like non-human primates, human infants’ vowel range is not large enough to accurately phonate the quantal vowels [a] and [i], which require the larger cavities of a two-tube bent vocal tract for sufficient supralaryngeal air flow; however, human infants are much better at phonating [u] because they are not impeded by the simian shelf (Lieberman 1975). Additionally, because human infants’ articulatory fine-control is somewhat more developed than that of nonhuman primates, they can actually raise the back of their tongue to create a “bent” single-tube vocal tract appearance, which increases their vowel space significantly (Goldstein 1980).

Although no significant sexually dimorphic variation can be detected in children’s vocal tract morphology until puberty, when males experience additional laryngeal descent, there actually seems to be some form of sexually dimorphic vocalization in boys and girls older than four years of age (Perry et al. 2001) – an age where an adult vocal tract configuration has already been fully obtained (Vorperian et al. 2009). Almost impossible to notice for a listener, these differences are in fact due to very small but widespread morphological differences between male and female children’s vocal tracts. Following is a list of these sexually dimorphic features for different measured anatomical variables (Vorperian et al. 2009:4-8):

- Total vocal tract length, oral cavity length, pharyngeal cavity length, and pharynx wall length are all greater in males than in females, as is growth rate
- Nasopharyngeal length’s growth rate is higher in females before late adolescence, and higher in males thereafter

- Oropharyngeal width is higher in males than in females during early childhood and late adolescence, but higher for females during early adolescence; the same is true for growth rate
- Anterior cavity length is higher in females than in males, as is growth rate

Overall, all variables show significant sexual dimorphism after twelve years of age except anterior cavity length, with variables being greater in males than females for all variables after this age (Vorperian et al. 2009). Differential growth rates between the sexes associated with adolescence might account for these trends, but their absence before adolescence means that prepubertal vocal tract sexual dimorphism must be caused by something else. A possible explanation could be differences in growth patterns associated with different tissues in male and female children; males tend to have very high proportions of somatic growth associated with most vocal tract variables, whereas females also have some significant neural growth – although most growth is still somatic in nature (Vorperian et al. 2009).

### **Vocal Tract Morphology and Speech Production in Fossil Hominins**

Ideally, the fossil record would be the best source for evidence concerning hominin speech and language emergence. In practice, however, this approach is marred by certain limitations that do not apply to those methodologies presented in the previous sections. As was mentioned before, the vocal tract consists exclusively of soft tissue – if the hyoid bone and laryngeal skeleton are excluded. As such, researchers have to rely on skeletal features to reconstruct hominins' vocal tracts, and such attempts are problematic at best and outright wrong at worst, due to the large amount of conjecture that is inherent to the

process (Lieberman 1975). Early studies analyzing fossil evidence for speech frequently were troubled by such falsely-drawn conclusions because they focused exclusively on individual morphological features while neglecting those features' integrated greater framework; they did not necessarily view features from a purely functional perspective; they lacked comparative data, such as from extant nonhuman primates; or, especially in the very early days of research that preceded the source-filter theory, they did not have a full understanding of how speech is produced (Lieberman 1975). And even if all these factors are kept in mind, the limited amount of data available for research – i.e., known fossils – complicates this approach immensely, a fate that is intrinsic to paleoanthropology as a whole.

That having been said, the fossil record remains of tremendous importance for any sort of insight into the emergence of speech. Obviously, any hominin is evolutionarily closer to modern humans than chimpanzees, let alone other extant primates. Additionally, knowing the state of vocal tract anatomy in a fossil taxon as well as its degree of relation to that of *Homo sapiens* makes it possible to consider a series of “stepping stones” from a nonhuman primate vocal tract configuration to one that closely resembles a modern one. Unfortunately, this model does not work out as nicely in reality, with “primitive” and “modern” morphology occasionally dis- and reappearing at various times in hominin evolutionary history, and with no clear connection to other trends of morphological change such as increased orthognathism (Lieberman 1975). However, a general trend of increasingly “modern” morphology does seem to be present, with australopithecine vocal tracts resembling chimpanzee vocal tracts more closely than later *Homo* vocal tracts (Lieberman 1975).

Out of all hominin taxa, the best-known one both to science and the general populace are Neanderthals (either *Homo neanderthalensis* or *Homo sapiens neanderthalensis*, depending on what taxonomy is used – the taxonomic status of Neanderthals would naturally have large implications on language capabilities, and this question is thus one that must be resolved for further consideration). This is true for their vocal tract anatomy as well, with most knowledge available coming from several cranial reconstructions over the course of the 20<sup>th</sup> century. Two of the earliest were made by Pierre Marcellin Boule in 1913 and 1921 (Boe et al. 2002), who reconstructed the well-known La Chapelle-aux-Saints specimen (La Chapelle-aux-Saints 1); the reconstruction that was most significant for research into Neanderthal vocalization was made by Lieberman & Crelin (1971), who based their reconstruction on Boule's, but put specific focus on vocal tract anatomy (Lieberman 1975; Boe et al. 2002). Lieberman & Crelin found that Neanderthal vocal tract morphology more closely resembled that of chimpanzees and, more specifically, modern human infants, which confirmed a number of earlier observations with similar results (Crelin 1969; Vlek 1979). They subsequently concluded that Neanderthals lacked the full range of modern human speech (Falk 1975) (Fig. 8). According to Lieberman & Crelin (1971:3-10), Neanderthals had:

- geniohyoid muscles that extended only posteriorly, and not also inferiorly as in modern humans; this indicates less descent of the hyoid bone
- an inward flaring mandible (hence the lack of a prominent chin, which is a hallmark morphological characteristic of *Homo sapiens*), and thus no accommodation for geniohyoid muscle attachment sites

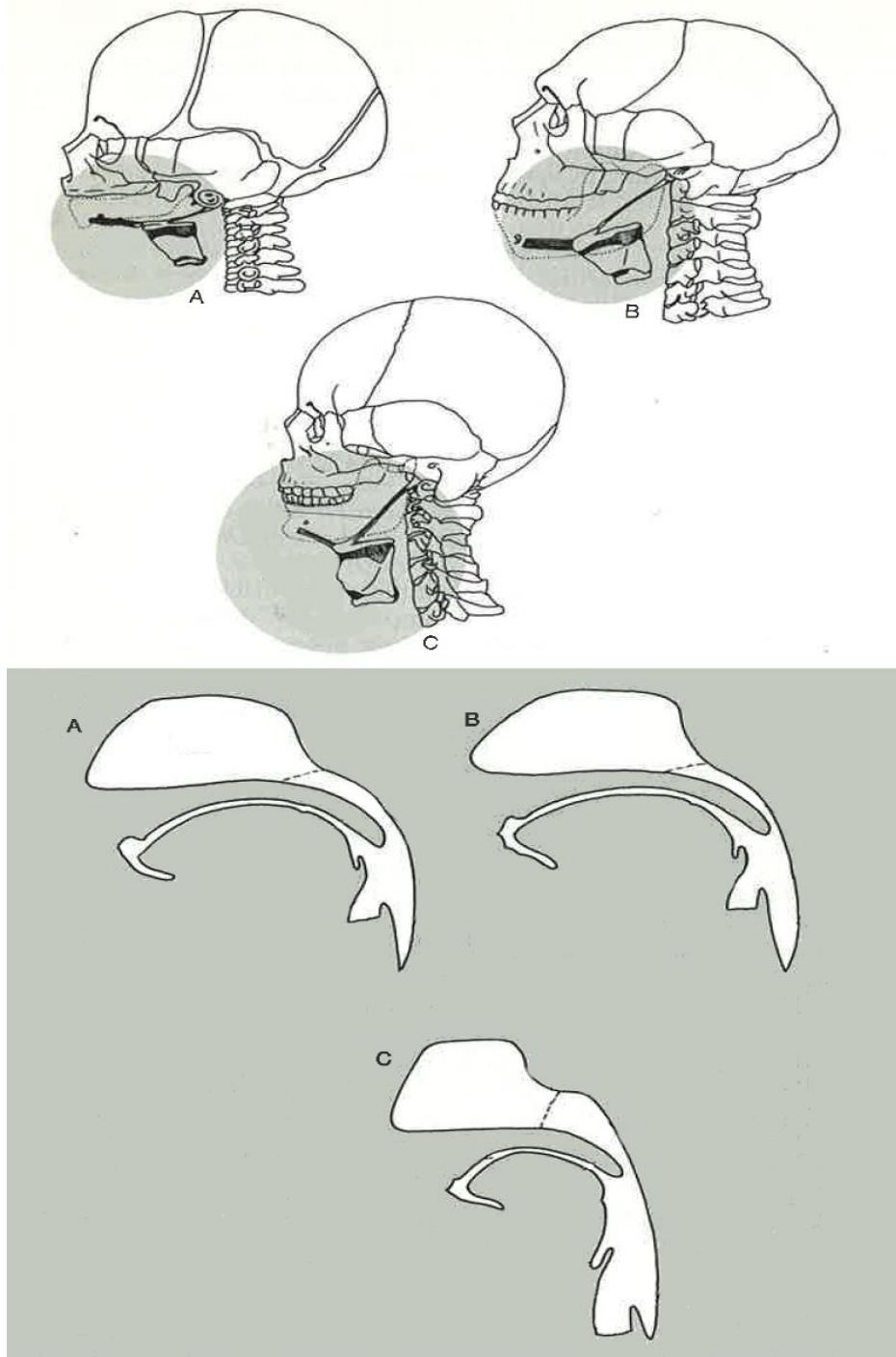


Fig. 8: *Crania and Reconstructed Laryngeal Structure (Above) and Reconstructed Vocal Tract (Below) for Modern Human Infants (A), Neanderthals (B), and Adult Modern Humans (C); modified from Lieberman (1975), after Lieberman & Crelin (1971)*

- more horizontally inclined styloid processes similar to those of modern human infants, indicating weaker stylohyoid muscles due to less favorable muscle-attachment sites
- a medial pterygoid plate that is more similar to that of modern human infants; this skeletal feature houses the muscle attachment sites for the pharyngeal constrictor muscles, which are involved in swallowing; in modern human adults, they also play a critical part in speaking
- a pushed-back sphenoid bone, similar to that of modern human infants
- an increased thoracic vertebral canal similar to that of adult modern humans, which indicates that respiratory fine-control had already emerged in Neanderthals
- no laryngeal air sacs, as indicated by the absence of hyoid bullae, which are walled shell-like bony structures that support the laryngeal air sacs in nonhuman primates; the bullae's presence was no longer necessary due to advanced respiratory fine-control
- a larger oral cavity than adult modern humans, and thus a >1:1 oral cavity to pharyngeal cavity ratio; unlike in modern human infants, the larger oral cavity of Neanderthals is caused not only by a relatively smaller pharyngeal cavity, but also by increased prognathism
- less descent of the larynx compared to adult modern humans, and a laryngeal position similar to that of modern human infants (La Chapelle-aux-Saints 1 was an elderly male; because laryngeal descent is present in both chimpanzees and modern humans, it is likely that this was also the case for Neanderthals, and thus a Neanderthal infant might have had an even higher larynx; also, the larynx of a

Neanderthal female would likely also have been higher, if a similar amount of sexual dimorphism for laryngeal descent was present in Neanderthals as in modern humans)

- less descent of the hyoid bone, in order to accommodate the more superiorly-positioned larynx
- a single-tube supralaryngeal vocal tract configuration, which is indicated by a number of features, such as a flattened basicranium and a tongue that was too large to accommodate a bent two-tube vocal tract – the tongue would have been stretched out, pushing the larynx all the way below the neck into the chest

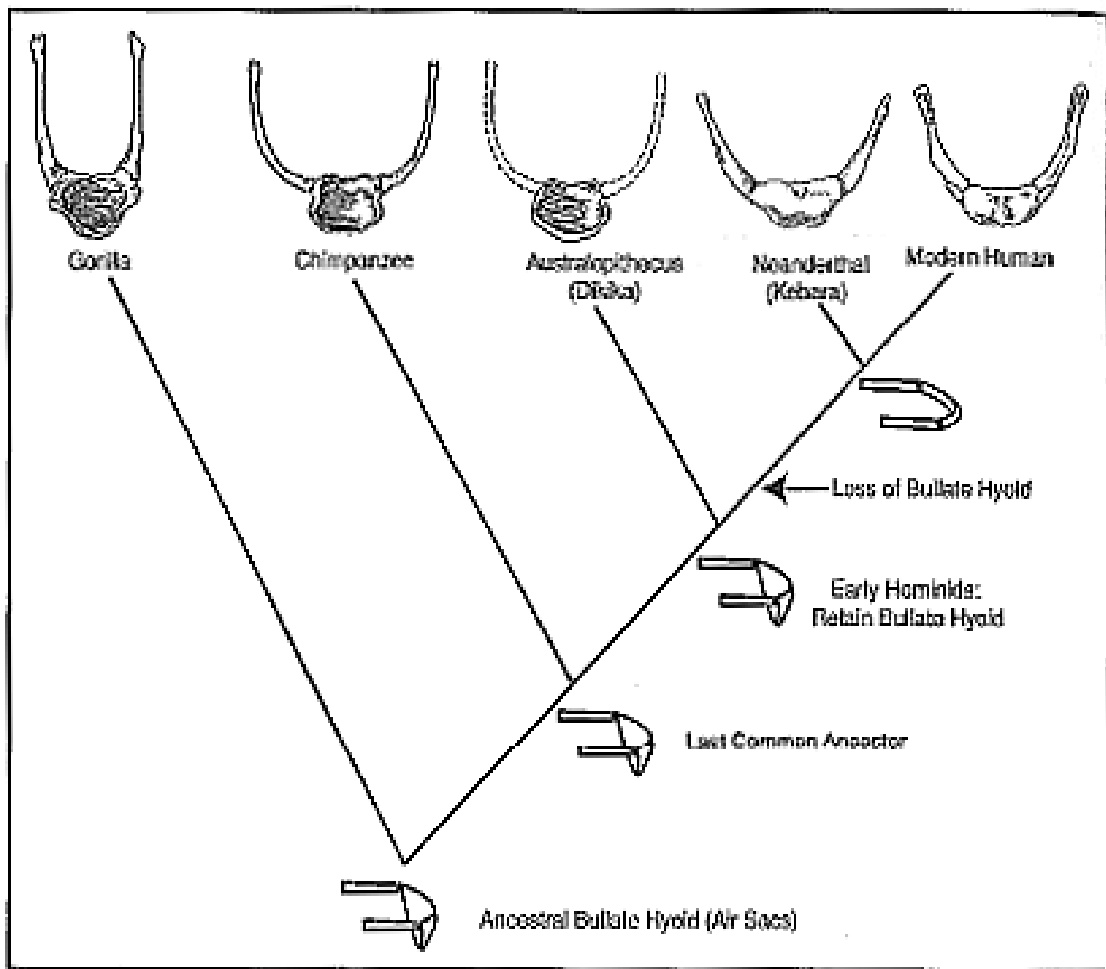
Based on these morphological results, Lieberman & Crelin postulated a number of hypotheses concerning Neanderthal vocalization and speech production. They assumed that, similar to nonhuman primates and modern human infants, Neanderthals could not distinctively phonate the quantal vowels [a] and [i]. Due to their larger oral cavity – and thus more room for expansion – coupled with the absence of a simian shelf, Neanderthals probably phonated [u] more clearly than modern human infants, but still not as precisely as modern human adults because they lacked the necessary bent two-tube supralaryngeal vocal tract configuration. This also means that Neanderthal vocalization was most likely exclusive nasal, similar to that of nonhuman primates; nasalization alone increases error rate on part of the listener by a minimum of 30% (Lieberman 1975), and thus would have seriously put Neanderthals at a disadvantage even if they otherwise had fully modern language – assuming Neanderthal speech perception would have been the same as in modern humans. Because nothing is known about speech perception as well as any other neural features of Neanderthals, it cannot be said for certain whether they had some sort

of language; but, based on vocal tract morphology alone, Lieberman & Crelin argued that any such language would not have been similar to that of modern humans and most likely less efficient as well. Following this assumption – if only spoken language is considered – it is entirely possible that Neanderthals were linguistically “outperformed” by more efficiently communicating *Homo sapiens* migrant populations in areas where there was population overlap, which would eventually have contributed to Neanderthal extinction (Lieberman 1991).

Paleoanthropology being paleoanthropology, Lieberman & Crelin’s results were challenged immediately. Critics pointed out that Lieberman & Crelin’s reconstruction underestimated laryngeal descent in Neanderthals by placing the hyoid bone at a position that was too high (Falk 1975). Lieberman & Crelin’s estimate of the hyoid bone position is based on its intersection with the stylohyoid ligaments and geniohyoid muscles, which are attached to the styloid processes; however, La Chapelle-aux-Saints-1 has very poorly preserved styloid processes, making any estimation reliant on these features spurious at best (Falk 1975). Additionally, the second, third, and fourth cervical vertebrae are missing (Straus & Cave 1957), making reconstruction of the hyoid bone based upon its position relative to these vertebrae impossible as well (Falk 1975). Complementing this is the conundrum of a 46° stylohyoid-ligament-to-zygomatic-arch angle coupled with a too-high hyoid bone making it impossible for Neanderthals to swallow similarly to modern humans due to conflicting muscle action and crowding, as well as mechanical lack of space in the pharyngeal cavity (Falk 1975). Besides these issues of a wrongly-placed hyoid bone, critics also argued that Lieberman & Crelin did not sufficiently take into account the great overall similarity in morphology of Neanderthal and modern human



hyoid bones (Arensburg et al. 1989), conforming to an overall morphological evolutionary trend of hyoid bone shape to a “modern” form (Fig. 9). However, in a statistical range of similarity for hyoid bone morphology, pigs’ hyoid bones are even more similar to those of modern humans than Neanderthal hyoid bones (Lieberman 1975). Seeing as pigs do not speak, assertions of hyoid bone morphology being essential for modern humanlike speech are thus somewhat problematic.



*Fig. 9: Hyoid Bone Morphology and Presence/Absence of hyoid bullae in (left to right) gorillas, chimpanzees, australopithecines, Neanderthals, and modern humans; Fitch (2010)*

More recent criticism of Lieberman & Crelin's findings has focused not on their reconstruction *per se*, but Boule's reconstruction on which it was based. Boule used a chimpanzee skeleton as a base model for his reconstruction, and more recent reconstructions by Jean-Louis Heim in 1986, 1989, and 1990 revealed the often-times gross errors that arose from this practice (Boe et al. 2002). Heim's (Heim 1990:100-113) reconstructions show – contrary to what Boule's reconstruction indicated – that:

- cranial positioning and basicranial angulation are the same in Neanderthals and adult modern humans
- no notable differences exist in laryngeal descent in Neanderthals and adult modern humans – an observation which was confirmed by the discovery of a Neanderthal cranium at Kebara, Israel, which had an articulated hyoid bone at a similar position as is the case in modern humans (Arensburg et al. 1989)
- palatal distance and oral cavity index – i.e., oral cavity height divided by palatal distance (Honda & Tiede 1998) – are very similar in adult modern humans and Neanderthals; these values also indicate a similar laryngeal position
- no significant differences can be assumed for hyoid-laryngeal musculature, due to overall similarities of styloid processes, basicranium, mastoid processes, and mandible; all of which can be used to estimate the position of the larynx and the dimension of the pharyngeal cavity (Laitman et al. 1979; Arensburg et al. 1985; Arensburg et al. 1989)

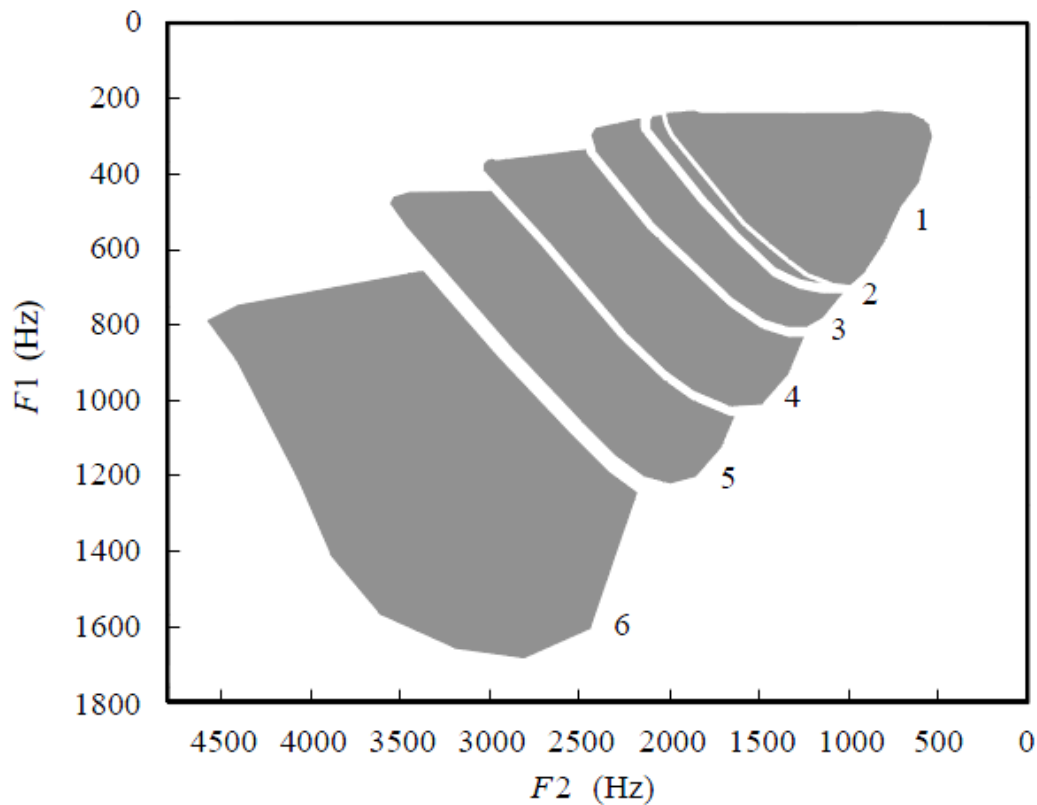
In addition to these findings, laryngeal height index (LHI) analysis for various modern human ethnic groups past and present, including, among others, ancient Egyptian and Andean mummies, as well as Neanderthals reveal that the latter comfortably fall within

the intragroup variation range of modern humans (Heim et al. 2000). Interestingly, however, Neanderthal values are closer to those of modern human females than modern human males – mostly due to the larger oral cavity in Neanderthals. The Neanderthal crania considered in this study – La Chapelle-aux-Saints-1 and La-Ferrassie-1 – both belong to adult males. A number of possible implications spring from this relationship. Assuming that Neanderthal laryngeal sexual dimorphism was similar to that of modern humans, the larynx of Neanderthal adult females and infants would accordingly have been positioned higher than that of adult males. Considering that the laryngeal height of adult modern human females is ideal for humanlike speech production, this would have made adult male Neanderthals more efficient speakers compared to Neanderthal females and infants; while adult female Neanderthals probably still fell into the range that is acceptable for efficient humanlike vocalization (de Boer 2010), Neanderthal infants might have been seriously compromised in their speech production based on laryngeal height. This was probably not a large disadvantage, as modern human infants only begin to speak once their initial laryngeal descent is complete, and a similar pattern can be assumed for Neanderthal infants. Alternatively, the >1:1 oral cavity to pharyngeal cavity ratio of Neanderthals might have favored a slightly higher larynx, making a hypothetical adult female Neanderthal larynx height the ideal for their specific vocal tract morphology. In adult male Neanderthals, then, just as in adult male modern humans, larynx height would be a compromise between efficient vocalization and size exaggeration. All these assumption, of course, only hold true if larynx height alone is considered, while all other vocal tract morphological differences are disregarded; and if similar patterns of laryngeal sexual dimorphism exist in Neanderthals and modern

humans. Unfortunately, because there is no evidence for adult female and infant Neanderthal vocal tract morphology as of yet, these assumptions cannot be tested.

Another important implication about Neanderthal speech capability must be made in comparison to human infants. Unlike nonhuman primates, human infants – who like nonhuman primates as well as Neanderthals have a larger oral cavity than pharyngeal cavity – are capable of raising the back of their tongues due to their more-developed tongue musculature (Goldstein 1980), which elongates the pharyngeal cavity; by doing so, human infants can somewhat equalize their oral cavity to pharyngeal cavity ratio. Neanderthals, whose oral musculature was probably similarly well-developed as that of modern humans as indicated by the presence of corresponding muscle-attachment sites, should have been capable of doing the same (Boe et al. 2002). Accordingly, Neanderthals could compensate for the deficiencies of their single-tube supralaryngeal vocal tract configuration by means of vocal articulation. As a consequence, it is very likely that Neanderthals' vowel space would have been only marginally smaller than that of adult modern human males and also larger than that of modern human infants due to stronger musculature and generally larger cavities (Boe et al. 2002) (Fig. 10). Neanderthals' vowel range would have been significantly smaller than that of adult modern human females, however, due to the relationship of larynx height and oral cavity to pharyngeal cavity ratio – while larynx height is the same in adult male Neanderthals and adult modern human females, the latter's ~1:1 oral-cavity to pharyngeal cavity gives them the edge in terms of maximum vowel space.

It should be noted again at this point that skeletal reconstructions able to only convey parts of the whole picture; for example, Neanderthals might still have retained the ability



*Fig. 10: Potential Maximum Vowel Space in Terms of F1 and F2 for an adult modern human male (1), an adult male Neanderthal (2), an adult modern human female or a 16-year old modern human male (3), a 10-year old modern human child (4), a 4-year old modern human child (5), and a newborn modern human infant (6); Boe et al. (2002)*

to temporarily descend their larynx during vocalization, like chimpanzees and many other mammals do (Fitch 2010). If this were the case, it is more than likely that Neanderthals were able to reduce the differences between theirs and modern humans' vowel space even further. However, since no other evidence of extinct hominins' speech production capabilities is currently available, research mostly needs to rely on those clues obtainable from the fossil record.

Compared to Neanderthals, relatively little evidence is available for other hominin species, due to the relatively small number of known fossils as well as worse preservation for the much older remains of australopithecines and early *Homo*. Nevertheless, some information can be extracted from what data is available to make a number of interpretations of early hominin vocalization. Lieberman & Crelin (1971), in addition to their reconstruction of the La Chapelle-aux-Saints-1 specimen, also reconstructed a number of crania from earlier hominin species, including the Sterkfontein-5 specimen (“Mrs. Ples”; *Australopithecus africanus*), the KNM-ER 3733 specimen from Koobi Fora, Kenya (*Homo ergaster/erectus*), the Broken Hill specimen (“Rhodesian Man”; *Homo heidelbergensis*), the Steinheim specimen (*Homo heidelbergensis*), the Jebel-Qafzeh-6 specimen (early *Homo sapiens*), and the Es-Skhul-5 specimen (early *Homo sapiens*).

Unlike their La Chapelle-aux-Saints-1 reconstruction, which was based on Boule’s earlier one, these reconstructions are original ones; criteria considered for the reconstructive process include known skeletal homologies of extinct and extant hominin species, muscle attachment sites, and comparative anatomy between hominins and nonhuman primates. For all reconstructions, the attempt was made to slightly skew them in order to bear greater resemblance to modern humans – the goal being to combat bias and keep results conservative.

Based on the Sterkfontein-5 cranial reconstruction, *Australopithecus africanus*’s vocal tract resembles that of chimpanzees and other nonhuman primates, but is relatively larger (Lieberman & Crelin 1971; Lieberman 1975). A similar vocal tract configuration can be assumed for the genus *Australopithecus* as a whole, and consequently, australopithecines’

vocal capabilities should closely correspond to those of chimpanzees, with the same limitations as far as modern humanlike speech is concerned (Lieberman 1975). Further evidence for this assumption comes from the fact that australopithecine and chimpanzee basicranial morphology also bears overwhelming resemblance (Lieberman 1991). However, while no morphological adaptations for speech seem to be present in australopithecines, their significantly larger brain sizes compared to chimpanzees supports the notion that non-speech-based protolanguage might have preceded speech in the evolution of the hominin lineage. Instead of vocalization, a hypothetical australopithecine protolanguage might have relied on other media of communication, such as gesturing (Hewes 1973). Alternatively, if australopithecines were capable of vocalization-linked temporal laryngeal descent – which is possible, given that some primates can do this – vocal protolanguage might have been entirely feasible in australopithecines (Fitch 2002). The permanently descended larynx that is found in modern humans might have consequently evolved to make speech faster, more precise, and more energy-efficient than speech relying on a temporally-descended larynx (Fitch 2002; de Boer 2010).

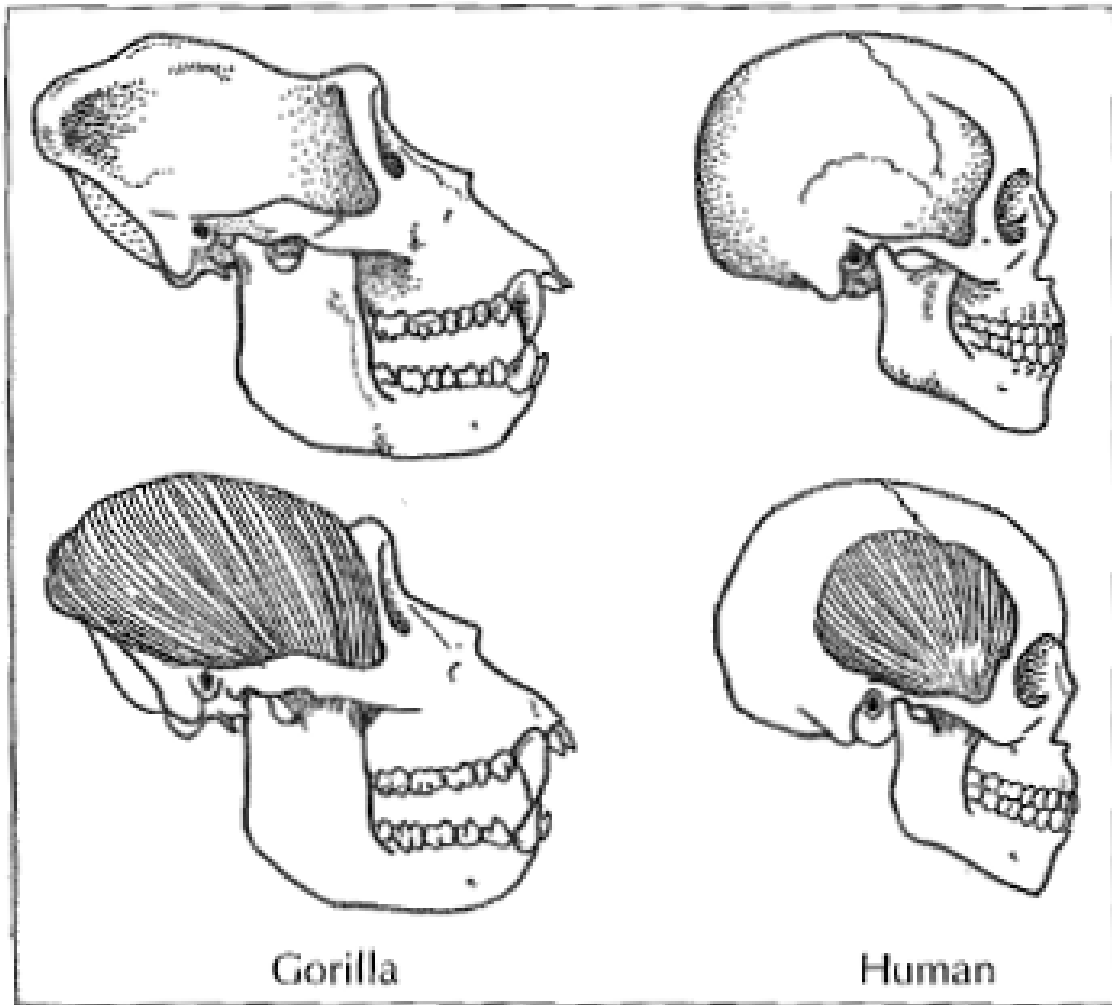
It is necessary to make some distinction between the gracile australopithecines and the robust australopithecines – the latter of which are sometimes grouped into their own genus called *Paranthropus*. Specifically, the immense differences in facial musculature, which is strongly pronounced in robust australopithecines such as *Australopithecus robustus* or *Australopithecus boisei*, could significantly have impacted speech production, as modern humanlike speech relies on orofacial muscular fine-control for precise articulation (Enard et al. 2002). The general consensus among

paleoanthropologists is that the strongly-developed facial musculature of robust australopithecines evolved for more efficient chewing (Reichholf 1997), and is unrelated to vocalization; however, it is possible that hyper-robust facial musculature of this kind actually hinders humanlike speech production. The vast reduction of the temporalis muscle that coincides with the emergence of genus *Homo* (Fitch 2010) (Fig. 11) probably occurred for dietary reasons, but it could also very well have been a necessary prerequisite for the much more recent emergence of orofacial muscular fine-control (Lieberman 2002), which is linked with the stabilization of the FOXP2 gene ~200,000 years ago (kya) (Enard et al. 2002) – coinciding with the emergence of *Homo sapiens*.

Just as is the case in terms of supralaryngeal vocal tract morphology, australopithecines also differ from *Homo* in that they retained laryngeal air sacs, as inferred from the presence of hyoid bullae (Alemseged et al. 2006). Complementarily, the thoracic vertebral canal of *Australopithecus afarensis* and *Australopithecus africanus* lacks the expansion that can be observed in modern humans and Neanderthals, implying that enhanced respiratory fine-control was absent in australopithecines (MacLarnon & Hewitt 1999). Thus, because of an inability to maintain the necessary air pressure for elongated expiratory phases, any hypothetical australopithecine protolanguage that relied on vocal communication would have been limited to short and relatively un-modulated utterances, similar to those of nonhuman primates (MacLarnon & Hewitt 1999).

Interestingly, thoracic vertebral expansion is also absent in early *Homo*, including *Homo habilis* and *Homo ergaster/erectus* (MacLarnon & Hewitt 1999). As a consequence, while early *Homo* might have had protolanguage that went significantly beyond australopithecine vocalization in other aspects, no innovations had yet taken place





*Fig. 11: Extent of the Temporalis Muscles in Western Lowland Gorillas (Gorilla gorilla) and modern humans; Fitch (2010)*

as far as enhanced respiratory fine-control is concerned. Supporters of the presence of language in early *Homo* argue that enhanced respiratory fine-control has been linked to other adaptations found in modern humans, such as bipedalism – the efficiency of which depends on upright maintenance of the trunk through thoracic and abdominal muscular action (Lovejoy 1988). Additionally, enhanced respiratory fine-control, while not essential, makes the kind of long-distance endurance running that is associated with

*Homo ergaster/erectus* subsistence strategies much more efficient (Carrier 1984). The evident lack of respiratory fine-control in early *Homo*, however, makes this hypothesis unlikely. Alternatively, increasingly difficult childbirth due to greater infant cranial size has been suggested to have necessitated greater respiratory – more specifically, the associated muscular – fine-control (Rosenberg 1992). While this issue was not yet present in *Homo habilis*, it definitely had arisen by the time of *Homo ergaster/erectus* (Begun & Walker 1993), which of course still lacked respiratory fine-control; also, the main muscle involved in childbirth is the uterine muscle, which is innervated from a different source than the respiratory muscles (MacLarnon & Hewitt 1999). In light of all these facts, it can be assumed that the main cause for enhanced respiratory fine-control by means of increased thoracic innervation is its necessity for complex modern humanlike speech production, and its absence in early *Homo* makes the presence of such speech unlikely (MacLarnon & Hewitt 1999; Ghazanfar & Rendall 2008).

While no differences exist between *Australopithecus* and early *Homo* in terms of respiratory fine-control, the situation is different in terms of laryngeal descent. The flexure of the basicranium in *Homo ergaster/erectus* indicates a permanently lower positioning of the larynx compared to australopithecines (Lieberman 1991) – an adaptation that very well might not initially have happened for vocalization purposes, but instead in order to facilitate enhanced breathing through the oral cavity in lieu of a lack of respiratory fine-control. Alternatively, permanent laryngeal descent might have occurred for size-exaggeration purposes (Nishimura et al. 2006; Fitch 2010), while a suggested adaption to bipedalism (Falk 1975; Aiello 1996) is unlikely due to the absence of a permanently descended larynx in australopithecines (Lieberman 1984).

Laryngeal descent is consecutively more pronounced in *Homo heidelbergensis*; no proportional relationship exists to increased orthognathism – i.e., palatal reduction – which in the Broken Hill cranium is relatively less developed than in the Steinheim cranium (Lieberman & Crelin 1971). At the same time, however, the major change that led to the establishment of a modern humanlike ~1:1 oral cavity to pharyngeal cavity ratio was increased orthognathism and not pharyngeal cavity expansion (Nishimura et al. 2006), and this overall trend of relatively less pharyngeal expansion can be observed in the Broken Hill specimen as well (Lieberman & Crelin 1971). As such, the exact role that increased orthognathism played in the emergence of a modern humanlike supralaryngeal vocal tract ratio remains on the table; however, as was already alluded to previously, the absence of modern humanlike phonation in certain orthognathic animals such as various toy breeds of dogs and cats, as well as hyper-orthognathic human infants suggests that it should not have been the main cause (Fitch 2010).

Although *Homo heidelbergensis* preceded Neanderthals in evolutionary time – and might have been the parent species of both Neanderthals and modern humans – their supralaryngeal vocal tract actually resembles that of modern humans more closely than that of Neanderthals (Lieberman 1975). While the Steinheim cranium's poor state of preservation makes any such inferences impossible, the Broken Hill specimen's skeletal morphology – such as the presence of a modern basicranium – suggests that *Homo heidelbergensis*'s supralaryngeal vocal tract was intermediate between a single-tube and a two-tube configuration in that it had the typical 90° bent angle also found in modern humans, but otherwise no clear boundary between oral and pharyngeal cavity (Lieberman & Crelin 1971). In terms of speech production capabilities, this means that – based on

skeletal clues alone – *Homo heidelbergensis* should have been able to clearly distinguish between the quantal vowels [a], [i], and [u] in their vocalization, but with less acoustic stability than is the case in modern humans (Lieberman 1975). Additionally, *Homo heidelbergensis* should have been able to produce un-nasalized speech, which Neanderthals were unable to do (Lieberman 1991). As a consequence from these morphological adaptations, *Homo heidelbergensis*'s skeletal morphology brought about impairments in terms of breathing, chewing, and swallowing, similar to those found in modern humans; the retainment and even further acceleration of these morphological features in modern humans point to the presence of enhanced vocal ability and speech production as an evolutionary driving factor (Lieberman 1991).

An essentially modern humanlike vocal tract configuration and all its associated features – a bent two-tube supralaryngeal vocal tract, pronounced permanent laryngeal descent, and enhanced respiratory fine-control – are definitely present in early *Homo sapiens*, as seen in the Jebel-Qafzeh-6 and Es-Skhul-5 specimens (Lieberman 1991). Accordingly, the overall integrated assortment of skeletal adaptations pertaining to modern human language seems to be species-specific (Lieberman 1975), and probably arose together with *Homo sapiens* as a distinct species. Whether language was present in early *Homo sapiens* cannot be entirely deduced from this fact because neural prerequisites play just as important a role in this process; however, the explosion of cultural expression left in the archeological record of the Middle and Upper Paleolithic – cave art, cultural artifacts, etc. – is a strong indicator that language was fully present in its current modern state by then at the latest (Lieberman 1975), and likely even before. A hypothetical hominin protolanguage that consisted of a mixture of gestures, mimics, and

vocalization might thus have become increasingly reliant on vocalization as the primary language component (Lieberman 1975), with the latter assuming the role it occupies today by 200-100kya.

The overall evolution of the vocal tract as seen in the fossil record emphasizes the primary role of speech in this process, as opposed to other factors such as bipedalism. Enhanced thoracic respiratory control arose sometime between 1.6mya and 100kya, as it is not present in the bipedal australopithecines and early *Homo*, but present in Neanderthals and modern humans (MacLarnon & Hewitt 1999). The same is true for pronounced laryngeal descent and a bent two-tube supralaryngeal vocal tract configuration, which arose much later than bipedalism (Lieberman 1975; Lieberman 1991). This assumption of speech being the main factor behind these morphological changes is further supported by modern medical cases of individuals that lack these adaptations; for example, individuals suffering from Apert's and Cruzon's syndrome which cause an overextension of the palate are unable to communicate effectively using speech, but show no impediments in terms of bipedal locomotion (Lieberman 1991). Because a compromised ability to communicate effectively would have put individuals at a selective disadvantage both in terms of natural as well as sexual selection (Lieberman 1991), and at the same time individuals with greater communicative ability would have been at a selective advantage due to being able to exchange potentially life-saving as well as mate-attracting information (Fitch 2010), and also using communication for social bonding purposes, it is likely that greater communicative control and articulatory ability gradually became entrenched as driving forces in the evolution of the vocal tract. The pace of this development – i.e., the increase in phonological transmission rates – was

gradually optimized by natural selection in that a balance was established between too-slow transmission – inefficient – and too-fast transmission – too much physiological stress on the brain and vocal apparatus (Desalles 2000). Human phonetic ability consequently accommodates this balance, leading to the modern relationship of communicative vocal expression and auditory perception that characterizes our species.

### **Conclusion**

Among the myriad of different animal taxa alive today, only humans use language as their main form of vocalization. However, as this paper shows, language did not grow out of nothing, but has its roots in the evolutionary past. Indeed, traces of the ability for communicative vocalization similar to human language can be observed both in living nonhuman primate and other mammal species as well as the fossil record; even in humans, such evidence is readily apparent in the maturation process. Language, then, is a biological feature as much as anything else, and as such it relies on a morphological framework for its proper functioning. The vocal tract constitutes this framework in humans, and together with the brain it is the most critical prerequisite for proper human language use.

In this paper, I approach the unique nature of the human vocal tract from four different perspectives. In section one, I discuss how each different part belonging to one of the three vocal tract subsystems works together to form an integrated framework, with each subsystem fulfilling a unique role for speech production. The way human language works is dependent on this morphology, and as such, the distinctive features of human speech as determined by vocal tract morphology are addressed as well.

To underline the unique nature of humanlike speech, I then turn to the vocalization of nonhuman primates in section two, discussing the differences in vocal tract anatomy between different taxa and what implications this has on vocalization behavior. By giving an overview of nonhuman primates' ability to emulate human speech, I raise the question whether the absence of a humanlike vocal tract prevents nonhuman primates from speaking – the answer being that while it may not be the only barrier, it definitely plays a large role.

In order to further emphasize that human vocal tract morphology is both evolutionarily derived as well as critical to human speech production, I look at vocal tract morphological development in human infants in section three. There, I argue that at birth, human infants' vocal tracts closely resemble those of nonhuman primates, and while some differences exist, human infants' vocalization is accordingly subject to the same limitations as far as speech production is concerned.

Lastly, in section four, I turn to the ultimate source for human vocal tract evolution – the fossil record. I emphasize the difficulties of extracting information from fossils, due to issues of preservation, fragmentary representation, and lack of context. At the same time, I discuss how findings concerning the language capabilities of extinct hominins have changed over time – specifically for Neanderthals. Nevertheless, a general trend from nonhuman primate-like towards more humanlike vocal tract morphology seems to be present in the fossil record. I take the stance that while earlier hominins might have had protolanguage or even some sort of rudimentary language, modern humanlike language as we know it is a species-specific innovation that arose in its present form together with *Homo sapiens*, based on differences in vocal tract morphology between

these taxa. As such, I do not rule out the possibility that extinct hominins like Neanderthals talked – but if they did, it would most likely have been substantially different from the way modern humans do.

Based on the results of this paper, some implications can be drawn. Most prominently, any hypothesis about language origin and evolution must keep vocal tract morphology in mind – while it may not be the only factor involved, its role is so important that it cannot be neglected either. Since human language is overwhelmingly speech-based – and perhaps was so to an even larger extent in the past, when writing was unknown and sign was the only viable alternative to spoken language – any morphological deficiency in the vocal tract would render humanlike language use exceedingly difficult, if not impossible. At the same time, however, other factors involved in human language must also be kept in mind. Specifically, neural input as far as language is concerned, as well as the evolution of the brain in general, are purposefully not addressed in this paper in order to maintain a reasonable scope of discussion. However, even if vocal tract morphology alone is considered, the brain plays a significant role by virtue of operating the vocal tract's functioning. As a result, future studies on this subject could explore the degree of neural input in speech production, as well as the evolutionary interrelationship between brain and vocal tract that lead to the one thing that makes humans essentially human – language.



## **Acknowledgements**

First and foremost, I would like to thank Dr. Mary J. Brody from the Department of Geography and Anthropology at LSU, for serving as thesis director and assisting me over the course of two semesters in writing this thesis. Without the numerous references and source materials she provided, as well as her advice on matters factual, technical, and administrative, this thesis would not have turned out the way it has.

I would also like to thank Dr. Robert Tague from LSU's Department of Geography and Anthropology, and Dr. Michael Hegarty from the Department of English for serving on the thesis defense committee for this thesis. Their critique and helpful comments have been much appreciated.

I would like to thank Marybeth Smith from the LSU Honors College as well. In her role as upper honors division and thesis coordinator, her guidance in how to organize an honors thesis, as well as her reminders of approaching deadlines were invaluable in keeping this project organized and on track.

Lastly, I would like to thank my parents for always being there for me when I needed it during this year in particular, my undergraduate college career in general, and throughout my life. *Ihr seid die Besten!*

## Bibliography

- Aiello, L.C.  
1996 Terrestriality, Bipedalism, and the Origin of Language. *Proceedings of the British Academy* 88: 269-289.
- Aiello, L.C., and R.I.M. Dunbar.  
1993 Neocortex Size, Group Size, and the Evolution of Language. *Current Anthropology* 34: 184-193.
- Aitchison, J.  
2000 *The Seeds of Speech*. Cambridge: Cambridge University Press.
- Alemseged, Z., Spoor, F., Kimbel, W.H., Bobe, R., Geraads, D., Reed, D., and J.G. Wynn.  
2006 A Juvenile Early Hominin Skeleton from Dikika, Ethiopia. *Nature* 443: 296-301.
- Andrew, R.J.  
1963 The Origin and Evolution of the Calls and Facial Expressions of the Primates. *Behaviour* 20: 1-109.  
1976 Use of Formants in the Grunts of Baboons and Other Nonhuman Primates. *Annals of the New York Academy of Sciences* 280: 673-693.
- Arensburg, B., Bar-Yosef, O., Chech, M., Goldberg, P., Laville, H., Meignen, L., Rak, Y., Tchernov, E., Tillier, A.-M., and B. Vandermeersch.  
1985 Une sepulture neandertalienne dans la grotte de Kebara (Israel). *Compte Rendu de l'Academie des Sciences de Paris* 300.2.6: 227-230.
- Arensburg, B., Tiller, A.M., Vandermeersch, B., Duda, H., Shephart, L.A., and Y. Rak.  
1989 A Middle Paleolithic Human Hyoid Bone. *Nature* 338: 758-760.
- Armstrong, D.F., Stokoe, W.C., and S.E. Wilcox.  
1994 Signs of the Origins of Syntax. *Current Anthropology* 35: 349-368.
- Barone, R.  
1976 *Anatomie comparee des mammiferes domestiques*. Paris: Vigot.
- Begun, D., and A. Walker.  
1993 The Endocast. In *The Nariokotome Homo erectus Skeleton*. Walker, A., and R. Leakey, eds. Pp. 326-358. Cambridge: Harvard University Press.
- Boe, L.-J., Heim, J.-L., Honda, K., and S. Maeda,  
2002 The Potential Neandertal Vowel Space Was as Large as That of Modern Humans. *Journal of Phonetics* 30: 465-484.
- Boe, L.-J., Heim, J.-L., Honda, K., Maeda, S., Badin, P., and C. Abry.  
2007 The Vocal Tract of Newborn Humans and Neanderthals: Acoustic Capabilities and Consequences for the Debate on the Origin of Language. *Journal of Phonetics* 35: 564-581.
- Bond, Z.S.  
1976 Identification of Vowels Excerpted from Neutral Nasal Contexts. *Journal of the Acoustical Society of America* 59: 1229-1232.
- Breuer, T., Ndoundou-Hockemba, M., and V. Fishlock.  
2005 First Observation of Tool Use in Wild Gorillas. *PLOS Biology* 3: e380.
- Broca, P.  
1861 Nouvelle Observation d'Aphemie Produite par une Lesion de la Moitie Posterieure des Deuxieme et Troisieme Ciconvolutions Frontales. *Bullet of the Paris Society for Anatomy* 6.2: 398-407.
- Brockelman, W.Y., and D. Schilling.  
1984 Inheritance of Stereotyped Gibbon Calls. *Nature* 312: 634-636.

- Calvin, W.H.  
 1992 Evolving Mixed-Media Messages and Grammatical Language: Secondary Uses of Neural Sequencing Machinery Needed for Ballistic Movements. *In* *Language Origin: A Multidisciplinary Approach*. Wind, J., Chiarelli, B., Bichakjian, B., and A. Nocentini, eds. Pp. 163-179. Dordrecht: Kluwer Academic Publishers.
- Campbell, E.J.M.  
 1968 The Respiratory Muscles. *Annals of the New York Academy of Sciences* 155: 135-140.
- Carre, R.  
 1996 Prediction of Vowel Systems Using a Deductive Approach. *Proceedings of ICSLP 96*: 434-437.  
 2004 From an Acoustic Tube to Speech Production. *Speech Communication* 42: 227-240.  
 2009 Dynamic Properties of an Acoustic Tube: Prediction of Vowel Systems. *Speech Communication* 51: 26-41.
- Carrier, D.  
 1984 The Energetic Paradox of Human Running and Hominid Evolution. *Current Anthropology* 25: 483-495.
- Chiba, T., and M. Kajiyama.  
 1958 The Vowel: Its Nature and Structure. Kyoto: Phonetic Society of Japan.
- Chomsky, N.  
 1972 Language and Mind. New York: Harcourt Brace Jovanovich.
- Conroy, G.C., and H. Pontzer.  
 2012 Reconstructing Human Origins. New York: W.W. Norton & Co.
- Crain, S.  
 1991 Language Acquisition in the Absence of Experience. *Behavioral and Brain Sciences* 14: 597-612.
- Crelin, E.S.  
 1969 Anatomy of the Newborn: An Atlas. Philadelphia: Lea & Febiger.  
 1987 The Human Vocal Tract. New York: Vantage Press.
- De Boer, B.  
 2009 Why Women Speak Better Than Men (and Its Significance for Evolution). *In* *The Prehistory of Language*. Botha, R., and C. Knight, eds. Pp. 255-265. Oxford: Oxford University Press.  
 2010 Modelling Vocal Anatomy's Significant Effect on Speech. *Journal of Evolutionary Psychology* 8: 351-366.
- De Boer, B., and W.T. Fitch.  
 2010 Computer Models of Vocal Tract Evolution: An Overview and Critique. *Adaptive Behavior* 18: 36-47.
- Deacon, T.W.  
 1997 The Symbolic Species. New York: W.W. Norton & Co.
- Desalles, J.-L.  
 2009 Why We Talk: The Evolutionary Origins of Language. Oxford: Oxford University Press.
- Diehl, R.L., Lindblom, B., Hoemeke, K.A., and R.P. Fahey.  
 1996 On Explaining Certain Male-Female Differences in the Phonetic Realization of Vowel Categories. *Journal of Phonetics* 24: 187-208.
- Draper, M.H., Ladefoged, P., and D. Whitteridge.  
 1959 Respiratory Muscles in Speech. *Journal of Speech, Language, and Hearing Research* 2: 16-27.
- DuBrul, E.L.  
 1958 Evolution of the Speech Apparatus. Springfield: Charles C. Thomas.
- Dunbar, R.I.M.  
 1993 Coevolution of Neocortical Size, Group Size and Language in Human Brains. *Behavioral Brain Science* 16L 681-735.

- Eldredge, N., and S.J. Gould.  
 1972 Punctuated Equilibria: An Alternative to Phyletic Gradualism. *In* Models of Paleobiology. Schopf, T.J.M., ed. San Francisco: Copper & Co.
- Enard, W., Przeworski, M., Fisher, S.E., Lai, C.S.L., Wiebe, V., Kitano, T., Monaco, A.P., and S. Pääbo.  
 2002 Molecular Evolution of FOXP2, a Gene Involved in Speech and Language. *Nature* 418: 869-872.
- Falk, D.  
 1975 Comparative Anatomy of the Larynx in Man and the Chimpanzee: Implications for Language in Neanderthal. *American Journal of Physical Anthropology* 43: 123-132.  
 1980 A Reanalysis of the South African Australopithecine Natural Endocasts. *American Journal of Physical Anthropology* 53: 525-539.
- Fant, G.  
 1960 Acoustic Theory of Speech Production. Berlin: De Gruyter Mouton.  
 1975 Non-Uniform Vowel Normalization. *Speech Transmission Laboratory Quarterly Progress and Status Report* 16: 1-19.
- Fitch, W.T.  
 1997 Vocal Tract Length and Formant Frequency Dispersion Correlate with Body Size in Rhesus Macaques. *Journal of the Acoustical Society of America* 102: 1213-1222.  
 2000a The Phonetic Potential of Nonhuman Vocal Tracts: Comparative Cineradiographic Observations of Vocalizing Animals. *Phonetica* 57: 205-218.  
 2000b The Evolution of Speech: A Comparative Review. *Trends in Cognitive Science* 4: 258-267.  
 2002 Comparative Vocal Production and the Evolution of Speech: Reinterpreting the Descent of the Larynx. *In* The Transition to Language. Wray, A., ed. Pp. 21-46. Oxford: Oxford University Press.  
 2010 The Evolution of Language. Cambridge: Cambridge University Press.
- 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 M.D. Hauser.  
 1995 Vocal Production in Nonhuman Primates: Acoustics, Physiology, and Functional Constraints on "Honest" Advertisement. *American Journal of Primatology* 37: 191-219.  
 2002 Unpacking "Honesty": Vertebrate Vocal Production and the Evolution of Acoustic Signals *In* Acoustic Communication. Simmons, A.M., Popper, A.N., and R.R. Fay, eds. Pp. 65-137. New York: Springer-Verlag.
- Fitch, W.T., and Kelley, J.P.  
 2000 Perception of Vocal Tract Resonances by Whooping Cranes (*Grus Americana*). *Ethology* 106: 559-574
- Fitch, W.T., and D. Reby.  
 2001 The Descended Larynx in Not Uniquely Human. *Proceedings of the Royal Society of London B: Biological Sciences* 268: 1669-1675.
- Flügel, C., and J.W. Rohen.  
 1991 The Craniofacial Proportions and Laryngeal Position in Monkeys and Man of Different Ages: A Morphometric Study Based on C-T Scans and Radiographs. *Mechanical Aging Development* 61: 65-83.
- Frey, R., and T. Riede.  
 2003 Sexual Dimorphism of the Larynx of the Mongolian Gazelle (*Procapra gutturosa pallas*, 1777) (Mammalia, Artiodactyla, Bovidae). *Zoologischer Anzeiger* 242: 33-62.
- Galdikas, B. M. F.  
 1982 An Unusual Instance of Tool-Use among Wild Orang-Utans in Tanjung Putting Reserve, Indonesian Borneo. *Primates* 23: 138-139
- Gardner, R.A., and B.T. Gardner.  
 1992 Early Signs of Language in Cross-Fostered Chimpanzees. *In* Language Origin: A Multidisciplinary Approach. Wind, J., Chiarelli, B., Bichakjian, B., and A. Nocentini, eds. Pp. 163-179. Dordrecht: Kluwer Academic Publishers.

- Geissmann, T.  
1984 Inheritance of Song Parameters in the Gibbon Song Analyzed in 2 Hybrid Gibbons (*Hylobates pileatus* s *H. lar*). *Folia Primatologica* 42: 216-225.
- Ghazanfar, A. A., and D. Rendall.  
2008 Evolution of Human Vocal Production. *Current Biology* 18.11
- Gilbert, H.R., Robb, M.P., and Y. Chen.  
1997 Formant Frequency Development: 15-36 Months. *Journal of Voice* 11: 260-266.
- Goldstein, U.G.  
1980 An Articulatory Model for the Vocal Tract of the Growing Children. Ph.D. dissertation, Massachusetts Institute of Technology.
- Goodall, J.  
1986 The Chimpanzees of Gombe: Patterns of Behavior. Cambridge: Harvard University Press.
- Green, S.  
1975 Variations of Vocal Pattern with Social Situation in the Japanese Monkey (*Macaca fuscata*): A Field Study. *In* Primate Behavior: Developments in Field and Laboratory Research 4. Rosenblum, L.A., ed. Pp. 1-102. New York: Academic Press
- Greenberg, J.  
1963 Universals of Language. Cambridge: Harvard University Press.
- Gray, H.  
1918 Anatomy of the Human Body. 20<sup>th</sup> edition. Lewis, W.H., ed. Philadelphia: Lea & Febiger.
- Hauser, M.D.  
1991 Sources of Acoustic Variation in Rhesus Macaque (*Macaca Mulatta*) Vocalizations. *Ethology* 89: 29-46.
- Hayama, S.  
1970 The *Saccus larynges* in Primates. *Journal of Anthropological Sciences of Nippon* 78: 274-298.
- Hayes, K.J., and C. Hayes.  
1951 The Intellectual Development of a Home-Raised Chimpanzee. *Proceedings of the American Philosophical Society* 95:105-109.
- Heim, J.-L.  
1990 La Nouvelle Reconstruction du Crane Neandertalien de la Chapelle-aux-Saints: Methode et Resultats. *Bulletin et Memoires de las Societe d'Anthropologie de Paris* 6: 94-117.
- Hewitt, G., MacLarnon, A., and K.E. Jones.  
2002 The Functions of Laryngeal Air Sac in Primates: A New Hypothesis. *Folia Primatologia* 73: 70-94.
- Hewes, G.W.  
1973 Primate Communication and the Gestural Origin of Language. *Current Anthropology* 14: 5-24.
- Hixon, T.J., Mead, J., and M.D. Goldman.  
1976 Dynamics of the Chest Wall during Speech Production: Function of the Thorax, Rib Cage, Diaphragm and Abdomen. *Journal of Speech, Language and Hearing Research* 19: 297-356.
- Hixon, T.J., and G. Weismer.  
1995 Perspectives on the Edinburgh Study of Speech Breathing. *Journal of Speech, Language and Hearing Research* 38: 42-60.
- Hoit, J.D., Banzett, R.B., Brown, R., and S.H. Loring.  
1990 Speech Breathing in Individuals with Cervical Spinal Cord Injury. *Journal of Speech, Language and Hearing Research* 33: 798-807.

- Hoit, J.D., Shea, S.A., and R.B. Banzett.  
1994 Speech Production during Mechanical Ventilation in Tracheostomized Individuals. *Journal of Speech, Language and Hearing Research* 37: 53-63.
- Holloway, R.L.  
1983 Human Palaeontological Evidence Relevant to Language Behaviour. *Human Neurobiology* 2:105-114.
- Honda, K., and M.K. Tiede.  
1998 An MRI Study on the Relationship between Oral Cavity Shape and Larynx Position. *Proceedings of the 5<sup>th</sup> International Conference of Spoken Language Processing* 2: 437-440.
- Houghton, P.  
1993 Neandertal Supralaryngeal Vocal Tract. *American Journal of Physical Anthropology* 90: 139-146.
- Isaac, G.L.  
1976 Stages of Cultural Elaboration in the Pleistocene: Possible Archeological Indicators of the Development of Language. *Annals of the New York Academy of Sciences* 280: 275-288.
- Itani, J.  
1963 Vocal Communication in the Wild Japanese Monkey. *Primates* 4: 11-66.
- Jakobson, R.  
1968 Child Language Aphasia and Phonological Universals. . Berlin: De Gruyter Mouton.
- Jordan, J.  
1971 Studies on the Structure of the Organ of Voice and Vocalization in the Chimpanzees. *Folia Morphologica (Warsz.)* 30: 97-126.
- Kelemen, G.  
1948 The Anatomical Basis of Phonation in the Chimpanzee. *Journal of Morphology* 82: 229-256.
- Kellogg, W.N.  
1968 Communication and Language in the Home-Raised Chimpanzee. *Science* 162: 423-427.
- Kent, R.D., and A.D. Murray.  
1982 Acoustic Features of Infant Vocalic Utterances at 3, 6, and 9 months. *Journal of the Acoustical Society of America* 72: 353-65.
- Kiang, N.Y.-S., and W.T. Peake.  
1960 Components of Electrical Responses Recorded from the Cochlea. *Ann Otol Rhinol Laryngol* 69: 448-458.
- Krebs, J.R., and R. Dawkins.  
1984 Animals Signals: Mind Reading and Manipulations. *In Behavioral Ecology*. Krebs, J.R., and N.B. Davies, eds. Pp. 380-402. Sunderland: Sinauer Associates.
- Ladefoged, P.  
1968 Linguistic Aspects of Respiratory Phenomena. *Annals of the New York Academy of Sciences* 155: 141-151.
- Ladefoged, P., and K. Johnson.  
2010 A Course in Phonetics. 6<sup>th</sup> Edition. Stamford: Cengage Learning.
- 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*. J. Bosma, ed. Pp. 206-219. Washington, D.C: U.S. Government Printing Office.
- Laitman, J.T., Heimbuch, R.C., 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: 323-343.

- Laitman, J.T., and Reidenberg, J.S.  
 1993 Specialization of the Human Upper Respiratory and Upper Digestive System as Seen through Comparative and Developmental Anatomy. *Dysphasia* 8: 318-325.
- Leakey, M.  
 1981 Olduvai Gorge. Cambridge: Cambridge University Press.
- Lenneberg, E.H.  
 1967 Biological Foundations of Language. Hoboken: Wiley.
- 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., McCarthy, R.C., Hiiemae, K.M., and J.B. Palmer.  
 2001 Ontogeny of Postnatal Hyoid and Larynx Descent in Humans. *Archives of Oral Biology* 46: 117-128.
- Lieberman, P.H.  
 1975 On the Origins of Language: An Introduction to the Evolution of Human Speech. New York: Macmillan.  
 1984 The Biology and Evolution of Language. Cambridge: Harvard University Press.  
 1991 Uniquely Human: The Evolution of Speech, Thought, and Selfless Behavior. Cambridge: Harvard University Press.
- Lieberman, P.H., Klatt, D.H., 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.H., and E.S. Crelin.  
 1971 On the Speech of Neanderthal Man. *Linguistic Inquiry* 2: 203-222.
- Lieberman, P.H., Laitman, J.T., Reidenberg, J.S., and P.J. Gannon.  
 1992 The Anatomy, Physiology, Acoustics and Perception of Speech: Essential Elements in Analysis of the Evolution of Human Speech. *Journal of Human Evolution* 23: 447-467.
- Liljencrants, J., and B. Lindblom.  
 1972 Numerical Simulation of Vowel Quality Systems: The Role of Perceptual Contrast. *Language* 48: 839-862.
- Lindblom, B.  
 1998 Systemic Constraints and Adaptive Change in the Formation of Sound Structure. In *Approaches to the Evolution of Language: Social and Cognitive Bases*. Hurford, J.R., Studdert-Kennedy, M., and C. Knight, eds. Pp. 242-265. Cambridge: Cambridge University Press.
- MacLarnon, A.M.  
 1987 Size Relationships of the Spinal Cord and Associated Skeleton in Primates. Ph.D. dissertation, University of London.  
 1993 The Vertebral Canal. In *The Nariokotome Homo Erectus Skeleton*. Walker, A., and R. Leakey, eds. Pp. 359-390. Cambridge: Harvard University Press.
- MacLarnon, A.M., and Hewitt, G.P.  
 1999 The Evolution of Human Speech: The Role of Enhanced Breathing Control. *American Journal of Physical Anthropology* 109: 341-363.
- MacNeilage, P.F.  
 1998 The Frame/Content Theory of Evolution of Speech Production. *Behavioral and Brain Sciences* 21: 499-546.  
 2008 The Origin of Speech. Oxford: Oxford University Press.
- Maddieson, I.  
 1984 Pattern of Sounds, 2<sup>nd</sup> Edition. Cambridge: Cambridge University Press.
- Mead, J., Bouhuys, A., and D.F. Proctor.  
 1968 Mechanisms Generating Subglottic Pressure. *Annals of the New York Academy of Sciences* 155: 177-182.

- Mehler, J., and E. Dupoux  
1990 Naitre Humain. Paris: Odile Jacob.
- Mitchell, H.I., Hoit, J.D., and P.J. Watson.  
1996 Cognitive Linguistic Demands for Speech Breathing. *Journal of Speech, Language, and Hearing Research* 39: 93-104.
- Moon, S.-J., and B. Lindblom.  
2003 Two Experiments on Oxygen Consumption during Speech Production: Vocal Effort and Speaking Tempo. *Proceedings of the 15<sup>th</sup> International Congress of the Phonetic Sciences*: 3129-3132.
- Müller, J.  
1848 The Physiology of the Senses, Voice, and Muscular Motion with the Mental Faculties. Trans. W. Baly. London: Taylor, Walton, and Maberly.
- 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 Publishing Company.
- Nishimura, T.  
2003 Comparative Morphology of the Hyo-Laryngeal Complex in Anthropoids: Two Steps in the Evolution of the Descent of the Larynx. *Primates* 44: 41-49.  
2005 Developmental Changes in the Shape of the Supralaryngeal Vocal Tract in Chimpanzees. *American Journal of Physical Anthropology* 126: 193-204.
- Nishimura, T., Mikami, A., Suzuki, J., and T. Matsuzawa.  
2003 Descent of the Larynx in Chimpanzee Infants. *Proceedings of the National Academy of Science, U.S.A.* 100: 6930-6933.  
2005 Development of the Laryngeal Air Sac in Chimpanzees. *International Journal of Primatology* 28: In Press.  
2006 Descent of the Hyoid in Chimpanzees: Evolution of Face Flattening and Speech. *Journal of Human Evolution* 51: 244-254.
- Noble, W., and I. Davidson.  
1991 The Evolutionary Emergence of Modern Human Behavior: Language and Its Archeology: *Man* 26: 223-253.
- Ohala, J.J.  
1984 An Ethological Perspective on Common Cross-Language Utilization of F0 of Voice. *Phonetica* 41: 1-16.
- Owren, M.J., Seyfarth, R.M., and D.L. Cheney.  
1997 The Acoustic Features of Vowel-like Grunt Calls in Chacma Baboons (*Papio cyncephalus ursinus*): Implications for Production Processes and Functions. *Journal of the Acoustical Society of America* 101: 2951-2963.
- Patterson, D.K., and I.M. Pepperberg.  
1998 Acoustic and Articulatory Correlates of Stop Consonants in a Parrot and Human Subject. *Journal of the Acoustical Society of America* 106: 491-505.
- Pepperberg, I.M.  
1999 *The Alex Studies: Cognitive and Communicative Abilities of Grey Parrots*: Cambridge: Harvard University Press.  
2005 An Avian Perspective on Language Evolution: Implications of Simultaneous Development of Vocal and Physical Object Combinations by a Grey Parrot (*Psittacus erithacus*). In *Language Origins*. Tallerman, M., ed. Pp.232-239. Oxford: Oxford University Press.
- Perry, T.L., Ohde, R.N., and D.H. Ashmead.  
2001 The Acoustic Bases for Gender Identification from Children's Voices. *Journal of the Acoustical Society of America* 109: 2988-98.
- Peterson, G.E., and H.L. Barney.  
1952 Control Methods Used in a Study of Vowels. *Journal of the Acoustic Society of America* 24: 175-184.



- Proctor, D.F.  
1974 Breathing Mechanics During Phonation and Singing. *In* Ventilatory and Phonatory Control Systems. Wyke, B., ed. Pp. 39-57. Oxford: Oxford University Press.
- Provine, R.R.  
1996 Laughter. *American Science* 84: 38-45.
- Puts, D.A., Gaulin, S.J.C., and K. Verdolini.  
2006 Dominance and the Evolution of Sexual Dimorphism in Human Voice Pitch. *Evolution and Human Behavior* 27: 283-296.
- Remez, R.E., Rubin, P.E., Pisoni, D.B., and T.D. Carell  
1981 Speech Perception Without Traditional Speech Cues. *Science* 212: 947-950.
- Redican, W.K.  
1975 Facial Expressions in Nonhuman Primates. *In* Primate Behavior: Developments in Field and Laboratory Research 4. Rosenblum, L.A., ed. Pp. 1-102. New York: Academic Press
- Reichholf, J.H.  
1997 Das Rätsel der Menschwerdung. Munich: Dtv.
- Reynolds, P.C.  
1972 Play, Language, and Human Evolution. Paper presented at 1972 meeting of the American Association for the Advancement of Science, Washington, D.C.
- Riede, T., and W.T. Fitch.  
1999 Vocal Tract Length and Acoustics of Vocalization in the Domestic Dog (*Canis familiaris*). *Journal of Experimental Biology* 202: 2859-2867.
- Riede, T., Bronson, E., Hatzikirou, H., and K. Zuberbuhler.  
2005 Vocal Production Mechanisms in a Non-human Primate: Morphological Data and a Model. *Journal of Human Evolution* 48: 85-96.
- Robb, M.P., Chen, Y., and H. Gilbert.  
1997 Developmental Aspects of Formant Frequency and Bandwidth in Infants and Toddlers. *Folia Phoniatrica et Logopedia* 49: 88-95.
- Rosenberg, K.R.  
1992 The Evolution of Modern Child-Birth. *Yearbook of Physical Anthropology* 35: 89-124.
- Sadler, T.W.  
2006 Medical Embryology. Philadelphia: Lippincott, Williams & Wilkins.
- Sasaki, C.T., Levine, P.A., Laitman, J.T., and E.S. Crelin.  
1977 Postnatal Descent of the Epiglottis in Man: A Preliminary Report. *Archives of Otolaryngology* 103: 169-171.
- Shubin, N.  
2008 Your Inner Fish: A Journey into the 3.5-Billion-Year History of the Human Body. New York: Vintage Books.
- Smith, D.R.R., Patterson, R.D., Turner, R., Kawahara, H., and T. Irino.  
2005 The Processing and Perception of Size Information in Speech Sounds. *Journal of the Acoustical Society of America* 117: 305-318.
- Sonntag, C.F.  
1921 The Comparative Anatomy of the Koala (*Phascolarctos cinereus*) and Vulpine Phalanger (*Trichosurus vulpecula*). *Proceedings of the Zoological Society of London* 39: 547-577.
- Stathopoulos, E.T., and C. Sapienza.  
1993 Respiratory and Laryngeal Function of Women and Men during Vocal Intensity Variation. *Journal of Speech, Language and Hearing Research* 36: 64-75.

- Stevens, K.N.  
1989 On the Quantal Nature of Speech. *Journal of Phonetics* 17: 3-45.
- Stevens, K.N., and A.S. House.  
1955 Development of a Quantitative Description of Vowel Articulation. *Journal of the Acoustic Society of America* 27: 484-493.
- Steklis, H.D., and M.J. Raleigh.  
1979 Requisites for Language: Interspecific and Evolutionary Aspects. *In Neurobiology of Social Communication in Primates: An Evolutionary Perspective*. Steklis, H.D., and M.J. Raleigh, eds. Pp. 283-314. New York: Academic Press.
- Straus, W.L., and Cave, A.J.E.  
1957 Pathology and the Posture of Neanderthal Man. *Quarterly Review of Biology* 32: 348-363.
- Takemoto, H.  
2001 Morphological Analyses of the Human Tongue Musculature for Three-Dimensional Modeling. *Journal of Speech, Language and Hearing Research* 44: 05-107.
- Titze, L.R.  
1989 Physiologic and Acoustic Differences between Male and Female Voices. *Journal of the Acoustical Society of America* 85: 1699-1707.  
1994 *Principles of Voice Production*. Englewood Cliffs: Prentice-Hall.
- Trubetzkoy, N.S.  
1939 *Principles of Phonology*. Trans. C. Baltaxe. Berkeley: University of California Press.
- Truby, H.M., Bosma, J.F., and J. Lind.  
1965 *Newborn Infant Cry*. Stockholm: Almqvist & Wiksell.
- Van den Berg, J.  
1958 Myoelastic-aerodynamic Theory of Voice Production. *Human Molecular Genetics* 11: 1161-1167.
- Van Dommelen, W.A.  
1993 Speaker Height and Weight Identification: A Reevaluation of Some Old Data. *Journal of Phonetics* 21: 337-341.
- Vlek, E.  
1970 Etude Comparative Onto-phylogenetique de l'Enfant due Pech-de-l'Aze par Rapport a d'autres Enfants Neandertaliens. *In L'Enfant due Pech-de-l'Aze*, D. Ferembach et al., eds. Pp. 139-186. Paris: Masson.
- Vorperian, H.K., Kent, R.D., Lindstrom, M.J., Kalina, C.M., Gentry, L.R., and B.S. Yandell.  
2005 Development of Vocal Tract Length during Early Childhood: A Magnetic Resonance Imaging Study. *Journal of the Acoustical Society of America* 117: 338-350.
- Vorperian, H.K., Wang, S., Chung, M.K., Schimek, E.M., Durtschi, R.B., Kent, R.D., Ziegert, A.J., and L.R. Gentry.  
2009 Anatomic Development of the Oral and Pharyngeal Portions of the Vocal Tract: An Imaging Study. *Journal of the Acoustical Society of America* 125: 1666-1678.
- Wall, C.E., and K.K. Smith.  
2001 Ingestion in Mammals. *In Encyclopedia of Life Sciences*. Group, N.P., ed. Pp. 1-6. London: MacMillan
- Warren, D.K., Patterson, D.K., and Pepperberg, I.M.  
1996 Mechanisms of American English Vowel Production in a Grey Parrot (*Psittacus erithacus*). *Auk* 113: 41-58.
- Weissengruber, G.E., Forstenpointner, G., Peters, G., Kubber-Heiss, A., and W.T. Fitch.  
2002 Hyoid Apparatus and Pharynx in the Lion (*Panthera leo*), Jaguar (*Panthera onca*), Tiger (*Panthera tigris*), Cheetah (*Acinonyxiubatus*) and Domestic Cat (*Felis silvestris f. catus*). *Journal of Anatomy* 201: 195-209.
- Winitz, H.  
1960 Spectrographic Investigation of Infant Vowels, *Journal of Genetic Psychology* 96: 171-181.

Winkworth, A.L., Davis, P.J., Adams, R.D., and E. Ellis.

1995 Breathing Patterns during Spontaneous Speech. *Journal of Speech, Language, and Hearing Research* 38: 124-144.

Wray, A.

2005 The Broadening Scope of Animal Communication Research. *In* *Language Origins*. Tallerman, M., ed. Pp.232-239. Oxford: Oxford University Press.

Zemlin, W.R.

1988 *Speech and Hearing Science: Anatomy & Physiology*, 3<sup>rd</sup> Edition. Englewood Cliffs: Prentice-Hall.