Comparative Vocal Production and the Evolution of Speech: Reinterpreting the Descent of the Larynx

W. Tecumseh Fitch
Harvard University
33 Kirkland St, Rm 982, Cambridge, MA 02138

(in press in: *The Transition to Language* (A. Wray, Ed.) Oxford: Oxford University Press.

Introduction

The Importance of Comparative Data

Since Darwin the comparative method has been recognized as one of the most powerful tools for deriving and testing hypotheses about evolution. The comparative method provides a principled way to use empirical data from living animals to deduce the behavioral abilities of extinct common ancestors, along with clues to the adaptive function of those abilities. For example, in the study of the evolution of speech, analysis of the vocal behavior of nonhuman primates can help identify homologies (characteristics shared by common descent), which in turn allow us to infer the presence or absence of particular characteristics in shared ancestors. Second, examples of convergent evolution (where similar traits have evolved independently in different lineages, presumably due to similar selective forces) can provide clues to the types of problems that particular morphological or behavioral mechanisms are "designed" to solve. Unfortunately, we know much more about human speech than we do about vocal communication in any other species, and thus the empirical database for comparative study is weaker than desirable. However, interest in nonhuman vocal production has exploded recently, in terms of both peripheral factors and central nervous control, and we can look forward to rapid advances in the near future.

The other major approach to analyzing the evolutionary history of a trait is based on the fossil record. Analyses of fossils can provide information about the order in which different traits were acquired during phylogeny (e.g. large brains and bipedalism in hominid evolution, or feathers and flight in birds), and thus place important constraints upon evolutionary hypotheses. Fossils can also provide dating information, thus allowing collation with data from other fossils and sites, and inferences about the climate, ecology and competitors faced by the fossilized organism when alive. Unfortunately, the fossil record is notoriously incomplete, and there are no guarantees that any individuals in key transitional stages in evolution will have been preserved. More importantly, fossils typically preserve only skeletal morphology, and thus provide reliable indicators of behavior only when the behaviors in question require unambiguous morphological modifications (as habitual bipedalism and powered flight do).

For the last thirty years, fossil analysis has been the dominant approach to the study of the evolution of speech. Paleoanthropologists have attempted to use the hominid fossil record to

deduce the timing and order of speech-related adaptations such as the descent of the larynx and enlargement of cortical regions. This approach was initiated by the seminal observation by Lieberman et al. (1969) that the human vocal tract differs from that in other primates in having a lowered larynx (Figure 1), a configuration that allows humans to make a wider variety of vowels that other primate species. Soon afterward, Lieberman and Crelin (1971) used a reconstruction of the Neanderthal soft tissue vocal tract, based on basicranial anatomy and some comparative data, to infer that larynx position in Neanderthals was closer to that of other primates than of modern humans. This suggested that Neanderthals could not make certain speech sounds that are typical of modern human languages. Although those authors never claimed that Neanderthals lacked language entirely, the paper spurred a vigorous (and still ongoing) debate about the speech and language capacities of Neanderthals, and extinct hominids in general (Falk 1975, DuBrul 1976, Arensburg et al. 1989, 1990). A review of this literature makes clear that there is still no general agreement about when articulate speech came to play the crucial role that it subserves in modern human language. This is not surprising, because the vocal tract is largely made up of soft tissue that does not fossilize, and thus there are no obvious skeletal indicators that would provide unambiguous evidence for speech. Thus, despite many years of hard work, new fossils, and creative new approaches to analysis, the currently available fossil data are inconclusive.

Meanwhile, the comparative approach to the evolution of speech has languished. Despite some early work on the anatomy and physiology of vocal production in nonhuman primates that adopted an explicitly comparative approach (Lieberman 1968, Lieberman et al. 1969), there has been very little further work in this vein until recently. In this paper I will review these recent comparative data on vocal production in nonhuman animals and explore their implications for theories of the evolution of the human vocal tract and speech capabilities. After some terminological preliminaries, I first review results from the comparative study of primate vocal communication, which provide crucial insights into the primitive features of primate communication systems that were inherited by our early hominid ancestors. I then turn to recent anatomical and physiological data on animal vocal production that demonstrates great mobility of the larynx in nonhuman mammals, and casting doubt upon the notion that skeletal features could provide unambiguous cues to vocal capabilities and suggesting alternative hypotheses for the descent of the larynx. I conclude that the comparative approach provides a rich source of hypotheses and predictions about vocal evolution, and in some cases the means for testing them empirically.

Terminology:

Language and Speech: "Language" is a system for representing and communicating complex conceptual structures, irrespective of modality. Theoretically, language might have originally

been encoded gesturally rather than vocally (Hewes 1973, Falk 1980, Corballis 1992, this volume). Signed languages and the written word are contemporary examples of non-spoken language. In contrast, "speech" refers to the particular auditory/vocal medium typically used by humans to convey language. Although speech and language are closely linked today, their evolution may not have been, and in any case their component mechanisms can be analyzed separately. The evolution of language presumably entailed complex conceptual structures, a drive to represent and communicate them, and systems of rules to encode them. The evolution of speech required vocalizations of adequate complexity to serve these linguistic needs, entailing a capacity for vocal imitation and learning, and a vocal tract with a wide phonetic range. The evolution of human speech may also have required perceptual specializations (Liberman & Mattingly 1985, Lieberman 1984, Hauser 1996), but there are currently no clear candidates for human-specific speech perception mechanisms, and in this paper I will focus mainly on speech production.

One of the great problems in studying the evolution of language from a comparative viewpoint is the lack of clear homologues to human language in the communication systems of other animals, most glaringly in other nonhuman primates. Fortunately, this limitation does not apply in the realm of speech. Most peripheral aspects of human vocal production are shared with other mammals, allowing us to analyze the evolution of speech by studying these aspects from a comparative perspective. The acoustics, musculature, innervation and peripheral motor control of human and animal vocal tracts are fundamentally similar, and are all open to experimental investigation. Such investigations have revealed a few key differences between human vocal production abilities and those underlying animal vocalizations, as detailed below.

Phonetics and Phonology: When discussing the evolution of speech, a key distinction is that between the mechanisms underlying human phonetic capabilities (the physiological capacity to make the wide variety of speech sounds used in contemporary human languages), and phonological capabilities (the neural capacities that underlie critical organizational structures of speech). While some researchers might consider the latter problem as part of speech evolution, others view it as an aspect of language evolution. This decision seems largely a matter of taste, as long as the distinction is made, and the topic at hand clearly defined. In the current review I focus exclusively on phonetic evolution, considering the evolution of phonology as an important problem in the evolution of language *per se*. For a Darwinian approach to phonology and the evolution of syllabic structure, see MacNeilage (1998).

0.1. Formants in Animal Communication

In this section I review recent work in the comparative vein that addresses the issue of potential evolutionary homologies between human speech and animal communication systems. I focus on the unquestioned importance of formants in human speech, and the much more sparse and recent data on the role of formants in animal communication. These new data unambiguously demonstrate that a wide variety of vertebrates, including many species of mammals and birds, both produce and perceive formants. This suggests that the use of formants in animal communication has a long evolutionary history, predating humans by millions of years, that can be readily explored with the comparative method.

Formants (from Latin *formare*, to shape) are the natural modes (or resonances) of the air contained in the vocal tract (Titze 1994, Lieberman & Blumstein 1988). The pattern and movements of formants provide the most important acoustic cues in human speech. Each individual formant acts like a bandpass filter, letting certain frequencies pass through it unchanged in amplitude, while damping out all other frequencies. A single formant is thus specified by its center frequency and a bandwidth. The formant center frequency (typically shortened to just "formant frequency") is the frequency that the formant allows to pass through with maximal amplitude. The bandwidth gives a measure of the breadth of this "spectral sweet spot". Perceptual experiments with humans indicate that formant bandwidth plays a relatively minor role in the perception of vowels and other formant-defined sounds, so most discussions of formants in human speech focus on their frequency. There is no non-technical word for the perceptual quality of formants, though it is generally one aspect of "timbre". Nonetheless, formants are quite salient: the difference between "beet" and "boot" spoken at one pitch, is a difference in formants alone (corresponding to the vowels /i/ and /u/). Note that there is no relationship between formants and voice pitch (which is determined by the he fundamental frequency, or vibration rate of the vocal "cords" in the larynx): their independence is one of the central tenets of modern speech science. Thus it is a serious but regrettably common error to confuse formants with pitch.

A single vocal tract has many formants, which together define the vocal tract transfer function: the complex, multi-peaked filter created by the combination of all formants. In general it is possible to measure at least six formants in a human transfer function, but only the bottom three are necessary for most human phonetic distinctions to be preserved. The overall pattern of these formants is critical in speech (as opposed to the frequency of any single formant). The single most important determinant of the transfer function is the length of the vocal tract that produces it: long vocal tracts produce low-frequency, densely-packed formant patterns. Thus all the formants of large men are shifted to lower frequencies relative to those of smaller men, or children (Fitch & Giedd 1999). Additionally, modifications in the overall shape of the vocal

tract modify the frequencies of the lower formants, shifting them from their default position in a simple tube. These time-varying changes in formant pattern provide the most fundamental acoustic cues in human speech. Synthesized signals that preserve the time-varying pattern of formant frequencies, but eliminate all other acoustic cues, are understandable as speech by most listeners (Remez et al. 1981).

0.1.1 Formants in Animal Vocalizations

Given the vast literature on formant production and perception in speech and singing (Fant 1960, Lieberman & Blumstein 1988, Sundberg 1987, and Titze 1994 provide reviews), the literature on formants in animal communication is extremely limited. However, recent comparative data provide a clear indication that formant frequencies are prominent in the vocalizations of many diverse vertebrate species, and suggest that they may be more important functionally than previously recognized. In this section I briefly review these data, and their implications for the ancestral use of formants in primate communication (see Fitch & Hauser, in press for more detail).

Lieberman (1968, Lieberman et al. 1969) was the first to analyze formants in the vocalizations of nonhuman primates, focusing mainly on the relatively restricted frequency range and lack of temporal variability of formants in primate calls. Except for a spectrographic analysis by Richman (1976), and some prescient comments by Andrew (1976), work on formants in primate communication ceased for almost 20 years. During that period, the work of Suthers and colleagues on bats (Suthers & Fattu 1973, Suthers et al. 1988) and birds (Suthers & Hector 1988, Suthers 1994) provided abundant documentation of the importance of vocal tract resonances in vocal production in those groups. Further exploration of the role of formants in birdsong was provided by Nowicki (1987) and Brittan-Powell et al. (1997).

The 1990's saw a renaissance of interest in the role of formants in primate vocal production. Hauser et al. (1993) used audio-video analysis to demonstrate a correlation between lip position and formants in rhesus macaques (as suggested by Lieberman 1968). Because lip protrusion lengthens the vocal tract, it lowers formants. Hauser and Schön Ybarra (1994) went on to demonstrate the causal link between the two by experimentally eliminating lip movements with xylocaine injections. Shipley et al (1991) described similar findings in domestic cats. Fitch & Hauser (1995) explored the significance of formants in the evolution of primate communication and their implications for "honest" cues to body size. The first unambiguous demonstration that spectral peaks in primate vocalizations represent formants was provided by Fitch's (1997) work demonstrating a correlation between formant frequencies and vocal tract length. This work also showed that formants correlate with body size in macaques, confirming that formants could provide an honest cue to body size (similar

correlations were found in dogs by Riede & Fitch 1999). Owren et al. (1997), in a study of formants in baboon grunts, verified that these calls have an extremely restricted formant space relative to human speech, as argued for macaques by Lieberman et al. (1969). In sum, this work makes it clear that formants play an important role in vocal production in a diverse set of primates, and suggests that formants could convey useful information about body size in mammals. Do other members of the same species perceive this information?

0.1.2 Formant Perception in Animals

There is a long history of using animals as subjects in speech perception experiments. In general, these studies indicate that, given appropriate training, animals can learn to use formants to discriminate among different human speech sounds. This ability has been demonstrated in both birds (Hienz et al. 1981; Dooling & Brown 1990, Dooling 1992) and mammals (baboons: Hienz & Brady 1988; dogs: Baru 1975; cats: Dewson 1964). However, the relevance of these studies to animal communication is limited by the facts that 1) human voice parameters were used, and these may not tap into the same auditory perceptual systems or resources that animals have evolved for perceiving their own vocalizations, and 2) that the animals received extensive training in these studies, and the results thus tell us less about preexisting auditory capabilities than about skills the animals can acquire. In addition, "talking" birds such as parrots and mynahs can imitate human formants (Klatt & Stefanski 1974; Nottebohm 1976; Warren et al. 1996) which implies an ability to perceive these cues in speech without training. Thus these studies show that many species have a latent ability to perceive formants, but not that this ability is used in their species-specific communication systems.

More conclusive evidence was provided by Owren's groundbreaking work with vervet monkeys. Owren and Bernacki (1988) used linear predictive coding (LPC) analysis of vervet "snake" and "eagle" alarm calls to separate characteristics of source waveform, presumed vocal tract filtering functions, and temporal patterning which distinguished these calls. Owren (1990) then used an operant paradigm to test classification of synthetic calls in which each of these characteristics was modified independently. The results indicated that changes in the spectrum played the dominant perceptual role in distinguishing the two call types, raising the possibility that changes in vocal tract shape conveys distinctive information in these calls. However, Owren did not isolate these spectral cues to ensure that they derived from formants rather than other sources (e.g. frication noise, laryngeal source, etc.). Another recent study (Sommers et al. 1992) compared the formant perception abilities of humans and macaques directly, and found that the monkeys perceived formants in synthesized signals as accurately as do humans. In both of these studies the use of an operant paradigm, with trained animal pressing levers in a laboratory setting, still allowed extensive learning to take place.

To overcome these difficulties, we developed (Fitch & Kelley 2000) a naturalistic testing paradigm to examine the spontaneous reaction of naive birds and primates to conspecific calls that varied only in formant frequency. Loudspeakers were hidden in bushes or behind blinds and thus invisible to the subjects. Using a habituation/dishabituation paradigm, a set of sounds were played repeatedly until the animal habituated (no longer showed a vigilant head-raise or head-turn response). Then, we played a resynthesized version of one of these calls that differed from the original only in its formant frequencies. These formant-shifted calls elicited a renewed response ("dishabituation") in most cases. A control playback showed that subjects were not responding to artifacts induced by synthesis. These studies demonstrate that nonhuman subjects, with no training, responded to changes in formant frequency in conspecific calls. Together, the findings reviewed here indicate that formant perception is not an evolutionary novelty evolved by humans, or by primates, and suggest that formants may play an important role in the communication systems of many non-human vertebrates.

0.1.3 The Evolution of Formant Perception

Understanding how formants came to assume their central role in human speech demands an understanding of the role they played in pre-linguistic hominids. Data from nonhuman mammals allow us to reconstruct several non-exclusive possibilities for the ancestral role of formants in acoustic communication. The first is that formants play a role in individual identification (Hauser 1992, Owren 1996, Rendall et al. 1996). Because each individual's vocal tract differs slightly in length, shape, nasal cavity dimensions and other anatomical features, differences in formant frequencies or bandwidths could provide cues to the identity of a vocalizer. Many vertebrates can distinguish the voices of different individuals, such as their offspring (or parents), or familiar and unfamiliar neighbors. Thus, a role for formants as a component of "vocal signatures" could be widespread in vertebrate communication systems.

Formants also provide an indication of the body size of a vocalizer. Vocal tract length is positively correlated with body size in humans, monkeys and dogs (Fitch 1997, Fitch & Giedd 1999, Riede & Fitch 1999). In turn, formant frequencies are closely tied to vocal tract length: large individuals with long vocal tracts have low formant frequencies. (This formant cue is completely independent from voice pitch, which in fact has no correlation with body size in adult humans; Künzel 1989). These correlations are surprisingly strong ($r = -0.88$ in macaques), suggesting that our mammalian ancestors could have used formant frequencies to accurately estimate body size from vocalizations. This would be useful in many contexts, e.g. gauging the size of a stranger in darkness or dense foliage. Perceptual mechanisms for judging size may in turn have provided a preadaptation for 'vocal tract normalization' (Fitch 1997), a critical aspect of speech perception whereby sounds from different-sized speakers are 'normalized' to yield equivalent percepts (Ladefoged & Broadbent 1957, Nearey 1978, Lieberman 1984). Such

normalization allows us to recognize the /i/ vowel of a child and an adult as "the same", despite significant differences in formant frequencies.

Summarizing this section, there is no doubt that the vocalizations of many different vertebrate species possess formants. It follows from the basic physics of vocal production and the anatomy of the vocal tract that formants can provide cues to body size, and/or to individual identity. Furthermore, perceptual studies have shown that many vertebrates perceive formants, both in human speech and in conspecific calls. Direct comparisons of humans and monkeys show that monkeys perceive formants as accurately as human beings. Thus, rather than being in any way specific to human spoken language, formants have a long evolutionary history among mammals, and perhaps all terrestrial vertebrates, and we can expect that human speech perception was built upon this fundamental, shared primitive basis. This suggests that perceptual mechanisms previously believed to be unique to humans and evolved for speech (e.g. vocal tract normalization, Lieberman 1984) may in fact represent much more ancient adaptations.

0.2 Vocal Tract Anatomy and Physiology in Humans and Animals

A central puzzle in the evolution of speech revolves around the fact that human vocal tract anatomy differs from that of other primates. Figure 1 shows MRI midsagittal sections through the heads of a chimpanzee and an adult human. It is evident that the human larynx rests much lower in the throat. Indeed, in most mammals, the larynx is located high enough in the throat to be engaged into the nasal passages, enabling simultaneous breathing and swallowing of fluids (Negus 1949, Crompton et al. 1997). This is also the case in human infants, who can suckle (orally) and breathe (nasally) simultaneously (Laitman & Reidenberg 1988). During human ontogeny, starting at about age three months, the larynx begins a slow descent to its lower adult position, which it reaches after three to four years (Sasaki et al. 1977, Senecail 1979, Lieberman et al. 2001). A second, smaller descent occurs in human males at puberty (Fant 1975, Fitch & Giedd 1999). A similar "descent of the larynx" presumably occurred over the course of human evolution. Until recently, a descended larynx was believed to be unique to humans.

Although this difference between humans and animals has been known for more than a century (Bowles 1889, Howes 1889), it wasn't until the late sixties that the acoustic significance of this configuration was recognized. Building upon advances in speech science, Lieberman and colleagues realized that the lowered larynx allows humans to produce a much wider range of formants than other mammals (Lieberman et al. 1969, 1972). The change in larynx position greatly expands our phonetic repertoire, because the human tongue can now move both vertically and horizontally within the vocal tract (this anatomical configuration, in the shape of an inverted L, is sometime termed a "two-tube" vocal tract). By independently

varying the area of the oral and pharyngeal tubes we can create a wide variety of vocal tract shapes and formant patterns. In contrast, a standard mammalian tongue rests flat in the long oral cavity, and cannot create vowels such as the /i/ in "beet" or the /u/ in "boot". Such vowels are highly distinctive, found in virtually all languages (Maddieson 1984), and play an important role in allowing rapid, efficient speech communication to take place. These vowels require extreme constriction in some vocal tract regions, and dilation in others, which the two-tube configuration allows.

Since these early observations, the descent of the human larynx has played a central role in discussions of the evolution of speech. The high larynx position in the Neanderthal vocal tract reconstructed by Crelin (Lieberman & Crelin 1971) suggested that Neanderthal speech abilities were limited, relative to modern humans. It was also suggested that nonhuman mammals (and Neanderthals) were unable to close the velum completely, and thus produced only nasalized (and thus less discriminable vowels). Despite significant controversy (Morris 1974, Falk 1975, DuBrul 1976, Arensburg et al. 1990) this idea has been perennially cited since then (e.g., Diamond 1992). Recent researchers have emphasized the descent of the human larynx even further, suggesting that changes in our vocal anatomy represented the crucial first step towards particulate speech (Studdert-Kennedy 1998), or even syntax (Carstairs-McCarthy 1998).

Unfortunately, these interesting ideas have always rested upon an inadequate comparative database. The work on animal vocal anatomy dates mostly from the 19th century, and was based exclusively upon dissections of dead animals. By the time new techniques allowing anatomical visualization of living animals were developed, such as x-ray film (cineradiography) or MRI, the study of comparative anatomy had fallen from favor. Although occasional critics pointed out that the crucial issue is not the resting position of the larynx, but its position during vocalization (Nottebohm 1976), it is only recently that modern imaging techniques have been applied to live vocalizing mammals (Fitch 2000a). These data indicate that mammals lower the larynx as a matter of course during vocalization, in some cases approaching the "two-tube" configuration typical of adult humans. Furthermore, new anatomical analyses show that humans are not unique in our laryngeal position, because several other species also have a permanently descended larynx. These new comparative data, reviewed below, suggest that the importance of the descent of the larynx in speech evolution has been overemphasized.

0.2.1 Cineradiographic Investigations of Animal Vocalization

Fitch (2000a) presented x-ray video observations of vocalization in four mammal species (dogs *Canis familiaris*, goats *Capra hircus*, pigs *Sus scrofa* and cotton-top tamarins, *Saguinus*

oedipus). The cineradiographic observations indicate that animal vocal tracts are surprisingly elastic and mobile, and that dead, formalin-fixed specimens provide a poor guide to the range of vocal movements available to the living animal. These data indicate that the vocal tract configuration of vocalizing animals, at least in dogs, pigs, goats and monkeys, is more similar to that of human talkers than was previously inferred on the basis of dissections of dead animals. In particular, all four nonhuman species examined can and do lower their larynges into the oral cavity during loud vocalizations, either to a relatively minor degree (goats) or to a surprisingly extensive degree (dogs). Finally, all four species appear to raise the velum, closing off the nasal airway, during loud vocalizations.

(Figure 2 about here)

Figure 2 provides x-ray video stills from Fitch (2000a), illustrating the typical vocal tract movements during vocalization (in this case, a dog barking). In the first frame, the larynx is in its resting position, high in the throat, and the epiglottis and velum are touching. Thus, the dog is breathing exclusively through its nose (as probably typical in mammals). In the second frame, the dog lowers the larynx considerably, raises the velum to close off the nasal cavity, and phonates. Thus the acoustic energy emanates from the oral cavity alone (the bark is not nasalized). This larynx is presumably lowered by contraction of the "strap" muscles that stretch from the breastbone up to the larynx and hyoid (the sternohyoid and sternothyroid muscles). After ceasing phonation, the larynx rises again to a high position, but the epiglottis and velum are not reengaged until the animal swallows. This sequence is typical of all of the mammals examined in the study, and thus may represent the typical mammalian vocal gesture.

The most likely reason for the retraction of the larynx during vocalization in these species is that sounds emitted through the mouth are louder. This is because the nasal cavity, with its complex coiled turbinates and large surface area, absorbs sound more than the oral cavity (Fitch 2000a). In this context, it should be noted that the animals in this study were also able to produce purely nasal vocalizations, with no lowering of the larynx. This intranasal vocal configuration appears to be typical of quiet vocalizations (e.g., pig grunts, dog whines), supporting the loudness hypothesis.

These data have important implications for the evolution of speech. First of all, the surprising mobility of the larynx during vocalization suggests that static anatomy provides a poor indication of physiological capability. Thus, attempts to divine larynx position from the skeletal remains of fossil hominids must be viewed with suspicion. If dogs can achieve a substantially lowered larynx, without any changes in basicranial angle or hyoid morphology, it seems likely that Neanderthals, other fossil hominids or chimpanzees could as well.

Second, the fact that other mammals assume a two-tube vocal tract configuration during vocalization, but nonetheless do not produce complex formant movements or articulate speech, suggests that the primary limitations on their vocal ability result from limits on neural signaling rather than of peripheral anatomy. Finally, these data suggest that a lowered larynx during vocalization is in fact a primitive trait that we share with other mammals, rather than a uniquely human adaptation. What is unusual about our species is that the adult human larynx is permanently lowered, rather than dropping only during vocalization. However, other recent data show that even this difference is not uniquely human.

0.2.2 Permanent Descent of Larynx in Nonhuman Mammals

Although the cineradiographic data indicate that the low position of the human larynx may not be as significant as previously supposed, this configuration is still unusual. But despite the claims of Negus (1949), this position is not unique to humans. In at least two species of deer, red deer *Cervus elaphus* and fallow deer *Dama dama*, the larynx of postpubertal males is permanently lowered to a resting position comparable to that in humans (Fitch & Reby, in press). Both of these species of deer produce loud, low-pitched roars during the autumn rut that serve to intimidate rival males (Clutton-Brock & Albon 1979) and to entice and perhaps accelerate ovulation in females (McComb 1987, 1990). During these roars, the larynx is withdrawn as far as physiologically possible, to the inlet of the thorax, thus exaggerating the lowering of the larynx that accompanies vocalization in other mammals to an almost ridiculous extreme. Although the rest position of the larynx is similar to that of adult humans, there are some differences. First, the velum in males is greatly elongated, which probably allows them to have velar/epiglottic contact during resting breathing. Second, the linkage between the larynx and the hyoid bone in these deer is a formed by a highly elastic and extensible thyrohyoid ligament. Thus, while the larynx descends, the hyoid bone stays high in its normal resting position in these deer (unlike in humans, or dogs, where the hyoid descends along with the larynx). These data make it quite clear that the "descent of the larynx" can no longer be considered a uniquely human trait, and indicate that there must be other possible adaptive reasons for laryngeal descent besides speech production.

It seems likely that some other species possess a similar ability to lower the larynx far beyond the normal mammalian level, or that of humans. The hyoid linkage of large cats of the genus *Panthera* (lions, tigers, leopards, jaguars) has been known to be quite unusual for more than a century, in that it contains an elastic ligament which allows the hyoid and larynx to descend far from the base of the skull (Owen 1834, Pocock 1916, Hast 1989). Although cineradiographic observations of roaring lions are obviously difficult to obtain and not yet available, it seems likely that this unusual hyoid anatomy allows them to lower the larynx during vocalizations, like the deer described above. In Arabian camels (*Camelus dromedarius*), the hyoid anatomy

is like that of red deer, with an elastic thyrohyoid ligament, suggesting that they too could lower the larynx during vocalization. Finally, the larynx occupies a permanently lowered position in koalas *Phascolarctos cinereus* (Sonntag 1921, Fitch, unpublished data). In koalas, the hyoid descends with the larynx, more closely matching the human situation than do deer. These observations, although preliminary, suggest that extreme laryngeal lowering might be more common in mammals than one would have previously guessed. However, much more extensive comparative investigations of vocal production in a wide range of living mammals will be necessary before a comprehensive comparative appraisal of mammalian vocal anatomy and physiology is possible.

To summarize this section, recent comparative data suggest that the descent of the larynx in humans, though undeniably real and significant, has been overemphasized. The cineradiographic observations show that the position of the mammalian larynx (along with the hyoid and tongue) in the vocal tract is quite flexible, and the current data suggest that, in general, mammals lower the larynx when producing loud vocalizations. These observations also call into the question the notion that static anatomy provides a valid indication of physiological (and thus phonetic) potential, and cast doubt on attempts to reconstruct vocal anatomy and speech capabilities on the basis of fossilized skeletal remains. In the next section I will briefly describe some of the more general implications of these data for the evolution of speech.

0.3 Implications of Comparative Data for the Evolution of Speech

0.3.1 The Phonetic Importance of the Descended Larynx

Although the observations above suggest that the descent of the larynx may not have played the crucial role in the evolution of humans posited by some theorists, the fact remains that the human larynx is unusual (though not unique) among mammals. Furthermore, the fundamental observation of Lieberman and colleagues (1969) still appears to hold true: humans produce a much wider range of formant frequencies than any other mammal species that has been analyzed (e.g., Owren et al. 1997). I suggest that a modified form of Lieberman's explanation for this fact is still the best explanation for the descended larynx in modern humans: the two-tube vocal tract allows us to produce wider range of vowels, and probably other speech sounds, than would a single-tube tract. I propose two modifications to Lieberman's original hypothesis which make it consonant with the new comparative data reviewed above.

The first proposed modification is that animal vocal tracts, with a high laryngeal resting position, do not provide an absolute anatomical barrier to a variety of vocal tract shapes, as was previously believed. Rather, with some muscular effort, a mammal such as a dog (or

chimp) can pull the larynx down into a human-like position, and then in principle have a much wider range of vocal tract shapes available. Indeed it seems likely that this is precisely how early hominids spoke: by retracting larynx as needed during vocalization, and then returning it to a resting intranarial position. Once the larynx is temporarily lowered, there is no obvious anatomical or physiological barrier to the kind of rapid articulatory movements characteristic of modern human speech. In seeking to understand why neither dogs or chimps in fact perform such movements, it appears that more attention will need to be paid in the future to the phonetic needs of their vocal communication system, and to the neural mechanisms that control the articulators. From this neuroethological viewpoint, there is an obvious gradualistic path from the temporarily descended larynx typical of mammals, to a permanently low larynx as in humans: for organisms which do a lot of talking (as we do) it may simply become more energetically efficient to leave the larynx low than to continually raise and lower it. A vocal tract that rests in the vocalization position may also aid in speed and accuracy of articulation, for example providing a solid and constant basis for the precise control necessary to make fricatives.

The second modification is that the current utility of the descended larynx (providing phonetic virtuosity) need not be identical to its original function at an earlier stage of phylogeny. Thus, some authors have proposed that the original descent of the larynx was an incidental byproduct of upright posture (DuBrul 1976), or of the facial retraction typical of hominid evolution. I do not personally find such arguments convincing (why don't other habitually upright mammals such as kangaroos or gibbons have a descended larynx? Why is the larynx still high in domestic animals bred to have short snouts, such as Pekingese dogs or Persian cats?). However, the form of the underlying argument seems sound: once the larynx attained a permanently low position, for whatever reason, the new vocal anatomy could be "exapted" for its phonetic utility. I will now explore another possible evolutionary route to a low larynx, that is consistent with available data on vocal anatomy and acoustics, and can also explain the presence of a low larynx in species like deer, which lack speech.

0.3.2 The Size Exaggeration Hypothesis for Descent of the Larynx.

An alternative hypothesis for the descent of the larynx takes as its starting point the fact mentioned earlier that formants are correlated with body size (Fitch 1997, Fitch & Giedd 1999, Riede & Fitch 2000). One effect of a lowered larynx is to increase vocal tract length (and consequently, to decrease formant frequencies). Thus, an animal with a lowered larynx can duplicate the vocalizations of a larger animal that lacks this ability, exaggerating the impression of size conveyed by its vocalizations. According to this "size exaggeration" hypothesis, the original selective advantage of laryngeal lowering was to exaggerate size and

had nothing to do with speech. This remains the sole function of the extreme laryngeal descent observed in male deer (and probably other mammals, like lions, as well). Although Ohala (1983, 1984) initially offered a similar proposal, focusing on human males, as a refutation of Lieberman's hypothesis, I suggest that the two are compatible, with size exaggeration providing a preadaptation for the evolution of speech. Once the larynx was lowered, the increased range of possible formant patterns was co-opted for use in speech. Consistent with the size exaggeration hypothesis, a second descent of the larynx occurs at puberty in humans, but only in males (Fitch & Giedd 1999). This second descent, at least, appears to be part of a suite of sexually-selected male pubertal changes that enhance apparent size, including shoulder broadening and facial hair growth.

The size exaggeration hypothesis is general, following from basic vocal anatomy and general acoustic principles, rather than being specific to hominids. This suggests that other taxa might show vocal tract elongation, and provide additional comparative insights into this phenomenon. A well-studied example is provided by a wide variety of bird species that possess an anatomical peculiarity called tracheal elongation. In these species, the trachea forms long loops or coils within the body. Because the bird sound source, called the syrinx, rests at the base of the trachea, this greatly elongates the bird's vocal tract, lowering its formant frequencies. A recent analysis (Fitch 1999) suggests that this serves to exaggerate the impression of size conveyed by vocalizations. Such exaggeration may be highly effective in animals that vocalize at night or from dense foliage. It is interesting to note that tracheal elongation in birds is not restricted to males. In species where both males and females are territorial, both sexes have elongated trachea, and in one species the trait is found only in females (Fitch 1999). These data suggest that Ohala's (1983) exclusive focus on laryngeal descent in human males, "to improve the male's ability to protect the family unit" (p. 13) was overly narrow. Given the role of females as primary caregivers in most primates, such a function might well be expected for a mother's voice as well.

If the size exaggeration hypothesis is correct, the descent of the human larynx is but one example of a pervasive phenomenon in evolution: convergent evolution of an anatomical mechanism to exaggerate their vocally-projected size, by diverse species. In humans, by hypothesis, this provided a necessary preadaptation which, together with important changes in neural control mechanisms, allowed hominids to more richly exploit the vocal domain for linguistic communication.

0.4 Conclusions

The comparative data reviewed in this paper make clear that the empirical study of living animals can provide a rich source of data, insights and testable hypotheses about the evolution

of human speech. Studies of animal formant production and perception have revealed that the most basic mechanisms underlying speech have a long evolutionary history, and suggest that certain perceptual mechanisms that were once believed uniquely human (e.g., vocal tract normalization) may in fact be part of the primitive perceptual toolkit inherited from our prelinguistic ancestors. Work on animal vocal production has shown that the descent of the larynx may also represent the exploitation of a pre-existing adaptation (for making loud sounds) present in many mammals. Finally, data from nonhuman species with elongated vocal tracts suggests that the initial impetus for the descent of the larynx in early hominids may have had nothing to do with speech, but instead functioned to exaggerate body size. In all of these cases, the comparative data allow us to identify and investigate examples of homology and analogy, and thus specify the starting point, selective forces and subsequent phylogenetic history of some critical components of modern human speech.

Despite this impressive start, there is still a dearth of comparative data relevant to speech evolution. One of the most pressing needs is for a better understanding of the neural control mechanisms that underlie our ability to produce and imitate speech sounds. The most prominent vocal ability that differentiates humans from other primates is vocal imitation. Humans, like many birds, seals, whales and dolphins, can imitate the sounds they hear, but our nearest primate relatives almost completely lack this ability (Studdert-Kennedy 1983, Janik & Slater 1997, Fitch 2000b). It is worth noting that Darwin (1871) appreciated the importance of imitation, an important component of his theory that adaptations for singing in early hominids provided a preadaptation for speech (pp. 56-62). Unfortunately, the neural basis of our imitative ability remains almost completely mysterious, even in humans. Given the obvious importance of this ability for the formation of large vocabularies (without which syntax would be valueless), we can only hope that more data on vocal imitation in humans and other mammals will become available.

The last few decades have seen an acceleration of theoretical speculation concerning the evolution of speech and language. Much of this theory is based upon scraps of evidence from fossil hominids. Given the limitations of these fossils, which are unable to strongly constrain hypotheses, or test predictions, about speech evolution, an obvious alternative source of data is the comparative study of living nonhuman species. However, besides the seminal work of Lieberman, and a few lone researchers like Hauser, Owren, Suthers and their colleagues, there has been surprisingly little modern research into mammalian vocal production, animal formant perception, or other topics extremely relevant to the evolution of speech. Almost all data on mammal vocal anatomy is fifty or more years old, and thus predates an adequate theory of vocal production. Thus aspects of vocal anatomy that can today be recognized as critical (e.g., descended larynges) went unnoticed or unmentioned by these anatomists. As a result, there is

still a great deal to learn about animal vocal production. Modern visualization and signal processing tools developed to study speech are only now beginning to be applied to nonhuman mammals, and with some 4000 species still unstudied, any sweeping generalizations about how mammals make and perceive sound are obviously premature. Nonetheless, the data reviewed here clearly illustrate the value of an empirical, comparative approach to understanding the evolutionary precursors of human speech perception and production.

Acknowledgments: The extremely helpful comments of Eric Nicolas, Alison Wray and an anonymous reviewer on an earlier version of this manuscript are gratefully acknowledged. Takeshi Nishimura, Kyoto Primate Research Institute, Japan, kindly provided the MRI image of the chimpanzee in Figure 1. Megan, the dog of A.W. Crompton, graciously barked for Figure 2.

References

Andrew, R. J. (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., Schepartz, L. A., Tillier, A. M., Vandermeersch, B., & Rak, Y. (1990). "A reappraisal of the anatomical basis for speech in middle Paleolithic hominids," American Journal of Physical Anthropology 83, 137-146.

Arensburg, B., Tillier, A. M., Vandermeersch, B., Duday, H., Schepartz, L. A., & Rak, Y. (1989). "A middle paleolithic human hyoid bone," Nature 338, 758-760.

Baru, A. V. (1975). "Discrimination of synthesized vowels [a] and [i] with varying parameters (fundamental frequency, intensity, duration and number of formants) in dog." In G. Fant & M. A. A. Tatham (Eds.), *Auditory Analysis and Perception of Speech* (Academic Press, New York).

Brittan-Powell, E. F., Dooling, R. J., Larsen, O. H., & Heaton, J. T. (1997). "Mechanisms of vocal production in budgerigars (*Melopsittacus undulatus*)," Journal of the Acoustical Society of America 101(1), 578-589.

Carstairs-McCarthy, A. (1998). "Synonymy avoidance, phonology and the origin of syntax." In J. R. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the Evolution of Language* (Cambridge University Press, New York).

Clutton-Brock, T. H., & Albon, S. D. (1979). "The roaring of red deer and the evolution of honest advertising," Behaviour 69, 145-170.

Corballis, M. (1992). "On the evolution of language and generativity," Cognition 44, 197-226.

Crompton, A. W., German, R. Z., & Thexton, A. J. (1997). "Mechanisms of swallowing and airway protection in infant mammals (*Sus domesticus* and *Macaca fasicularis*)," Journal of Zoology (London) 241, 89-102.

Dewson, J. H. (1964). "Speech sound discrimination by cats," Science 141, 555-556.

Diamond, J. (1992). *The Third Chimpanzee* (HarperCollins, New York).

Dooling, R. (1992). "Hearing in birds." In D. B. Webster, R. F. Fay, & A. N. Popper (Eds.), *The Evolutionary Biology of Hearing* (pp. 545-560). (Springer-Verlag, New York, NY).

Dooling, R. J., & Brown, S. D. (1990). "Speech perception by budgerigars (*Melopsittacus undulatus*): Spoken vowels," Perception & Psychophysics 47(6), 568-574.

DuBrul, E. L. (1976). "Biomechanics of speech sounds," Annals of the New York Academy of Sciences 280, 631-642.

Falk, D. (1975). "Comparative anatomy of the larynx in man and the chimpanzee: implications for language in Neanderthal," American Journal of Physical Anthropology 49, 171-178.

Falk, D. (1980). "Language, handedness, and primate brains: Did the Australopithecines sign?," American Anthropologist 82, 72-78.

Fant, G. (1960). *Acoustic Theory of Speech Production* (Mouton & Co., The Hague).

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(2), 1213-1222.
Fitch, W. T. (1999). "Acoustic exaggeration of size in birds by tracheal elongation: comparative and theoretical analyses," Journal of Zoology (London) 248, 31-49.
Fitch, W. T. (2000 b). "The evolution of speech: a comparative review," Trends in Cognitive Sciences.
Fitch, W. T. (2000 a). "The phonetic potential of nonhuman vocal tracts: Comparative cineradiographic observations of vocalizing animals," Phonetica.
Fitch, W. T., & Giedd, J. (1999). "Morphology and development of the human vocal tract: a study using magnetic resonance imaging," Journal of the Acoustical Society of America 106(3), 1511-1522.
Fitch, W. T., & Hauser, M. D. (1995). "Vocal production in nonhuman primates: Acoustics, physiology, and functional constraints on "honest" advertisement," American Journal of Primatology 37, 191-219.
Fitch, W. T., & Kelley, J. P. (2000). "Perception of vocal tract resonances by whooping cranes, *Grus americana*," Ethology 106(6), 559-574.
Hast, M. (1989). "The larynx of roaring and non-roaring cats," Journal of Anatomy 163, 117-21.
Hauser, M. D. (1992). "Articulatory and social factors influence the acoustic structure of rhesus monkey vocalizations: A learned mode of production?," Journal of the Acoustic Society of America 91, 2175-2179.
Hauser, M. D. (1996). *The evolution of communication* (MIT Press, Cambridge, MA).
Hauser, M. D., Evans, C. S., & Marler, P. (1993). "The role of articulation in the production of rhesus monkey (*Macaca mulatta*) vocalizations," Animal Behaviour 45, 423-433.
Hauser, M. D., & Schön Ybarra, M. (1994). "The role of lip configuration in monkey vocalizations: Experiments using xylocaine as a nerve block.," Brain and Language 46, 232-244.
Hewes, G. W. (1973). "Primate communication and the gestural origin of language," Current Anthropology 14, 5-24.
Hienz, R. D., & Brady, J. V. (1988). "The acquisition of vowel discriminations by nonhuman primates," Journal of the Acoustical Society of America 84(1), 186-194.
Hienz, R. D., Sachs, M. B., & Sinnott, J. M. (1981). "Discrimination of steady-state vowels by blackbirds and pigeons," Journal of the Acoustical Society of America 70(3), 699-706.
Janik, V. M., & Slater, P. B. (1997). "Vocal learning in mammals," Advances in the study of behavior 26, 59-99.
Künzel, H. J. (1989). "How well does average fundamental frequency correlate with speaker height and weight?," Phonetica 46, 117-125.
Laitman, J. T., & Reidenberg, J. S. (1988). "Advances in understanding the relationship between the skull base and larynx with comments on the origins of speech," Journal of Human Evolution 3, 99-109.
Liberman, A. M., & Mattingly, I. G. (1985). "The motor theory of speech perception revised," Cognition 21, 1-36.
Lieberman, P. (1968). "Primate vocalization and human linguistic ability," Journal of the Acoustical Society of America 44(6), 1574-1584.
Lieberman, P. (1984). *The Biology and Evolution of Language* (Harvard University Press, Cambridge, MA).
Lieberman, P., & Blumstein, S. E. (1988). *Speech Physiology, Speech Perception, and Acoustic Phonetics.* (Cambridge University Press, Cambridge, UK).
Lieberman, P., & Crelin, E. S. (1971). "On the speech of Neanderthal man," Linguistic Inquiry 2, 203-222.
Lieberman, P., Crelin, E. S., & Klatt, D. H. (1972). "Phonetic ability and related anatomy of the newborn and adult human, Neanderthal man, and the chimpanzee," American Anthropologist 74(3), 287-307.
Lieberman, P., Klatt, D. H., & Wilson, W. H. (1969). "Vocal tract limitations on the vowel repertoires of rhesus monkeys and other nonhuman primates," Science 164, 1185-1187.
MacNeilage, P. F. (1998). "The frame/content theory of evolution of speech production," Behavioral and Brain Sciences 21, 499-546.
Maddieson, I. 1984. *Patterns of sounds*. (Cambridge: Cambridge University Press).

McComb, K. (1987). "Roaring by red deer stags advances date of oestrous in hinds," Nature 330, 648-649.
McComb, K. E. (1991). "Female choice for high roaring rates in red deer, *Cervus elaphus*," Animal Behaviour 41, 79-88.
Nearey, T. (1978). *Phonetic Features for Vowels* (Indiana University Linguistics Club, Bloomington).
Nottebohm, F. (1976). "Vocal tract and brain: A search for evolutionary bottlenecks," Annals of the New York Academy of Sciences 280, 643-649.
Nowicki, S. (1987). "Vocal tract resonances in oscine bird sound production: evidence from birdsongs in a helium atmosphere.," Nature 325, 53-55.
Ohala, J. J. (1983). "Cross-language use of pitch: An ethological view.," Phonetica 40, 1-18.
Ohala, J. J. (1984). "An ethological perspective on common cross-language utilization of Fø of voice.," Phonetica 41, 1-16.
Owen, R. 1834. On the anatomy of the Cheetah, Felis jubata. *Transactions of the Zoological Society (London)* **1**, 129-136.
Owren, M. J. (1990). "Acoustic classification of alarm calls by vervet monkeys (*Cercopithecus aethiops*) and humans: II. Synthetic calls," Journal of Comparative Psychology 104, 29-40.
Owren, M. J. (1996). "An "acoustic-signature" model of speech evolution," Journal of the Acoustical Society of America 99, 2258.
Owren, M. J., & Bernacki, R. (1988). "The acoustic features of vervet monkey (*Cercopithecus aethiops*) alarm calls," Journal of the Acoustical Society of America 83, 1927-1935.
Owren, M. J., Seyfarth, R. M., & Cheney, D. L. (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.
Pocock, R. I. (1916). "On the hyoidean apparatus of the lion (F. leo) and related species of Felidae," The Annals and Magazine of Natural History 8(18), 222-229.
Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). "Speech Perception without traditional speech cues," Science 212, 947-950.
Rendall, D., Rodman, P. S., & Emond, R. E. (1996). "Vocal recognition of individuals and kin in free-ranging rhesus monkeys," Animal Behaviour 51, 1007-1015.
Richman, B. (1976). "Some vocal distinctive features used by gelada monkeys," Journal of the Acoustical Society of America 60, 718-724.
Riede, T., & Fitch, W. T. (1999). "Vocal tract length and acoustics of vocalization in the domestic dog *Canis familiaris*," Journal of Experimental Biology 202, 2859-2867.
Senecail, B. (1979). *L'Os Hyoide: Introduction Anatomique a l'Etude de certains Mecanismes de la Phonation* (Faculté de Médicine de Paris, Paris).
Sommers, M. S., Moody, D. B., Prosen, C. A., & Stebbins, W. C. (1992). "Formant frequency discrimination by Japanese macaques (*Macaca fuscata*)," Journal of the Acoustical Society of America 91(6), 3499-3510.
Studdert-Kennedy, M. (1998). "The particulate origins of language generativity: from syllable to gesture." In J. R. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the Evolution of Language* (Cambridge University Press, New York).
Sundberg, J. (1987). *The Science of the Singing Voice* (Northern Illinois University Press, Dekalb, Illinois).
Suthers, R. A. (1994). "Variable asymmetry and resonance in the avian vocal tract: A structural basis for individually distinct vocalizations," Journal of Comparative Physiology 175, 457-466.
Suthers, R. A., & Fattu, J. M. (1973). "Mechanisms of sound production in echolocating bats.," American Zoologist 13, 1215-1226.
Suthers, R. A., Hartley, D. J., & Wenstrup, J. J. (1988). "The acoustic role of tracheal chambers and nasal cavities in the production of sonar pulses by the horseshoe bat, *Rhilophus hildebrandti*.," Journal of Comparative Physiology, A 162, 799-813.
Suthers, R. A., & Hector, D. H. (1988). "Individual variation in vocal tract resonance may assist oilbirds in recognizing echoes of their own sonar clicks." In P. E. Nachtigall & P. W. B. Moore (Eds.), *Animal Sonar: Processes and Performances.* (pp. 87-91). (Plenum Press, New York).
Titze, I. R. (1994). *Principles of voice production* (Prentice Hall, Englewood Cliffs, N.J.).

Figure legends

Figure 1. Midsaggital MRI Images of Chimpanzee and Human: Tongue body is shaded. Note the reconfiguration of tongue shape in the adult human relative to the chimpanzee, and the low position of the larynx (behind tongue in chimp, beneath tongue in human)

Figure 2. Cineradiographic Still Images of Dog Barking: Left: Resting breathing is through nasal cavity, larynx is in standard, high position. Right: During barking, the larynx is retracted deep into the pharynx, drawing the tongue body along, and the velum raises to close of the nasal cavity. See Fitch (2000a) for details.

Figures

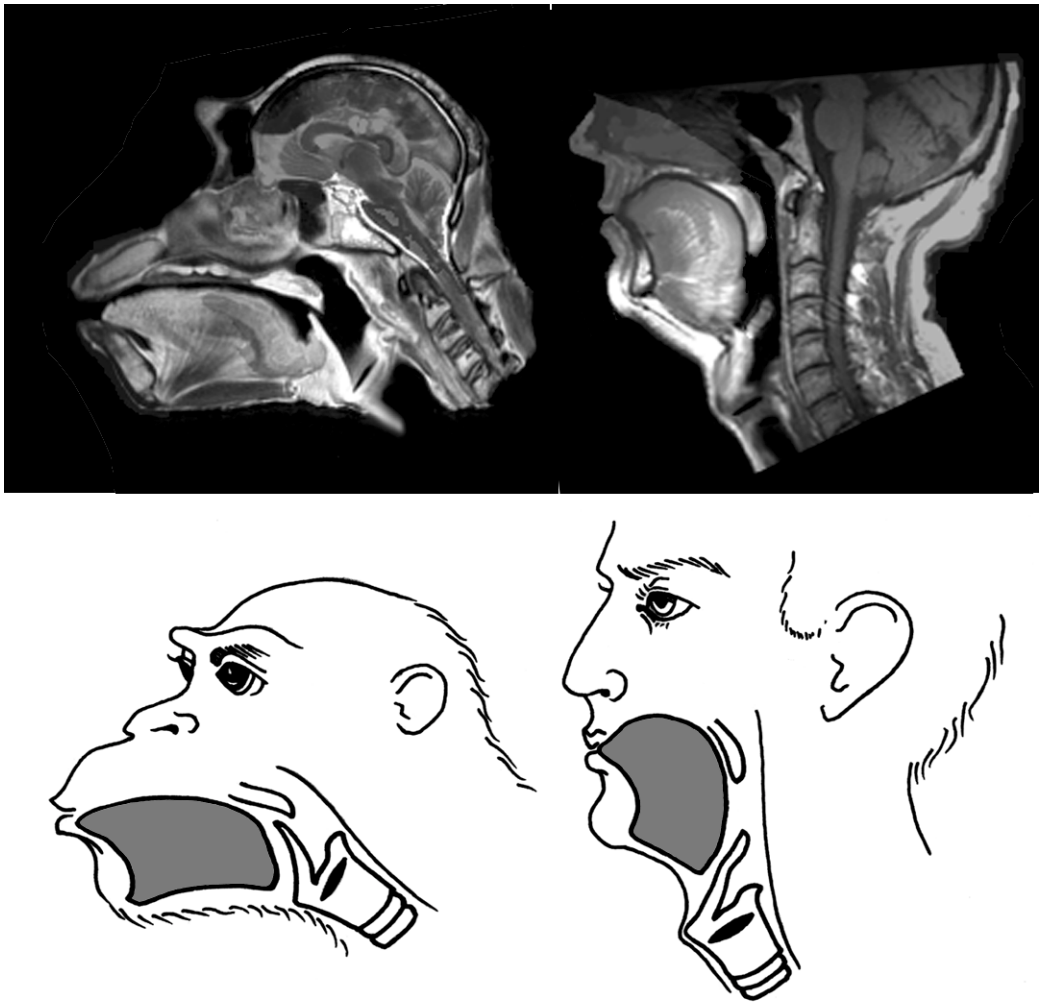

Figure 1

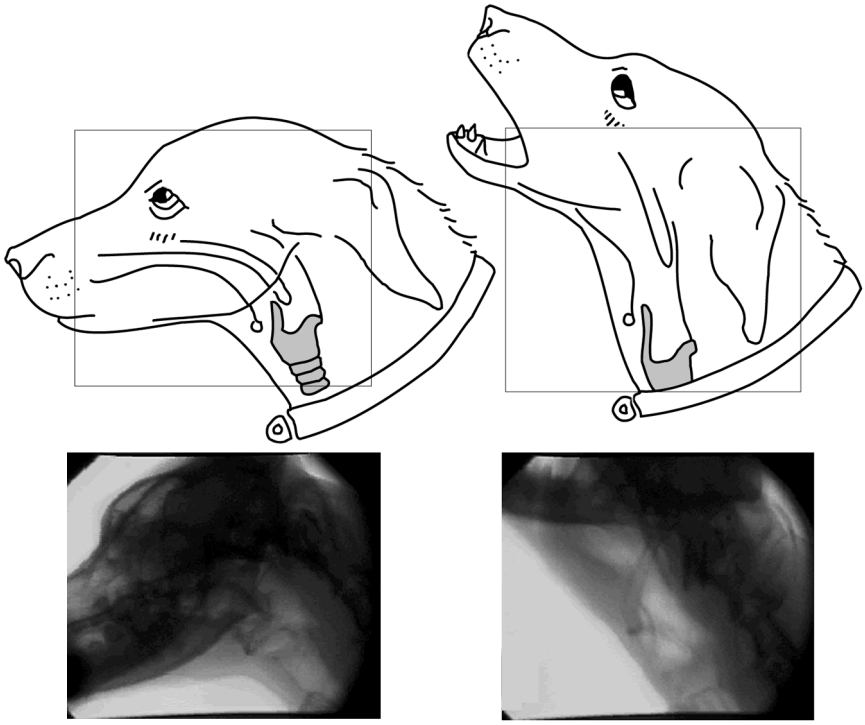

Figure 2