

# The evolution of speech: a comparative review

W. Tecumseh Fitch

**The evolution of speech can be studied independently of the evolution of language, with the advantage that most aspects of speech acoustics, physiology and neural control are shared with animals, and thus open to empirical investigation. At least two changes were necessary prerequisites for modern human speech abilities: (1) modification of vocal tract morphology, and (2) development of vocal imitative ability. Despite an extensive literature, attempts to pinpoint the timing of these changes using fossil data have proven inconclusive. However, recent comparative data from nonhuman primates have shed light on the ancestral use of formants (a crucial cue in human speech) to identify individuals and gauge body size. Second, comparative analysis of the diverse vertebrates that have evolved vocal imitation (humans, cetaceans, seals and birds) provides several distinct, testable hypotheses about the adaptive function of vocal mimicry. These developments suggest that, for understanding the evolution of speech, comparative analysis of living species provides a viable alternative to fossil data. However, the neural basis for vocal mimicry and for mimesis in general remains unknown.**

Most of us have heard talking birds such as parrots or mynahs. Some of us might have even heard Hoover, a harbor seal raised by a Maine fisherman who could convincingly (if a bit drunkenly) utter phrases like ‘Hey, hey you, get outta there’<sup>1</sup>. But none of us has ever heard a talking ape or monkey. This is not for lack of effort: even chimpanzees like Vicki, raised in diapers by the Hayes family, failed to obtain a rudimentary spoken vocabulary<sup>2</sup>. Nor is it due to a lack of communicative desire or ability: many independent experiments have demonstrated the ability of chimps trained with sign language or other visual symbol systems to develop large vocabularies and produce multi-symbol sentences. By contrast, their ability to imitate speech sounds consistently remained negligible<sup>3</sup>. Why do our nearest animal relatives lack vocal output capabilities that are comparable with ours? What changes needed to occur during human evolution before speech came to play its ubiquitous and irreplaceable role in human social interactions? What selective advantages drove the evolution of these abilities?

Although speech and language are sometimes treated as synonymous, it is important, in the evolutionary context, to distinguish them. ‘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<sup>4–6</sup>. Signed languages and the written word are contemporary examples of non-spoken language. By 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 component mechanisms can be analyzed separately. The evolution of language entailed complex conceptual structures, a drive to represent and communicate them, and systems of rules to encode them<sup>7</sup>. The evolution of speech required vocalizations of adequate complexity to serve linguistic needs, entailing a capacity for vocal learning, and a vocal tract with a wide phonetic range. The evolution of human speech might also have required perceptual specializations<sup>8,9</sup>, but this article focuses only on speech production.

The neural basis of language remains dimly understood, and homologies between language and animal communication systems are, at best, arguable. The absence of clear animal homologs to human language makes it difficult to study language evolution empirically. As a result, discussions of the evolution of language often involve more speculation than data, and the field has a checkered history (in 1866, the Linguistic Society of Paris banned all further discussion of the topic). By sharp contrast, most aspects of human vocal production are shared with other animals, which allows us to analyze the evolution of speech from a comparative evolutionary perspective. The acoustics, anatomy, innervation and central control of human and animal vocal tracts are fundamentally similar, and are amenable to experimental investigation. Such investigations have revealed a few key differences between human vocal production abilities and those that underlie animal vocalizations. This article describes these differences, and discusses the theory and data that have a bearing on how, when and why these differences arose during hominid evolution.

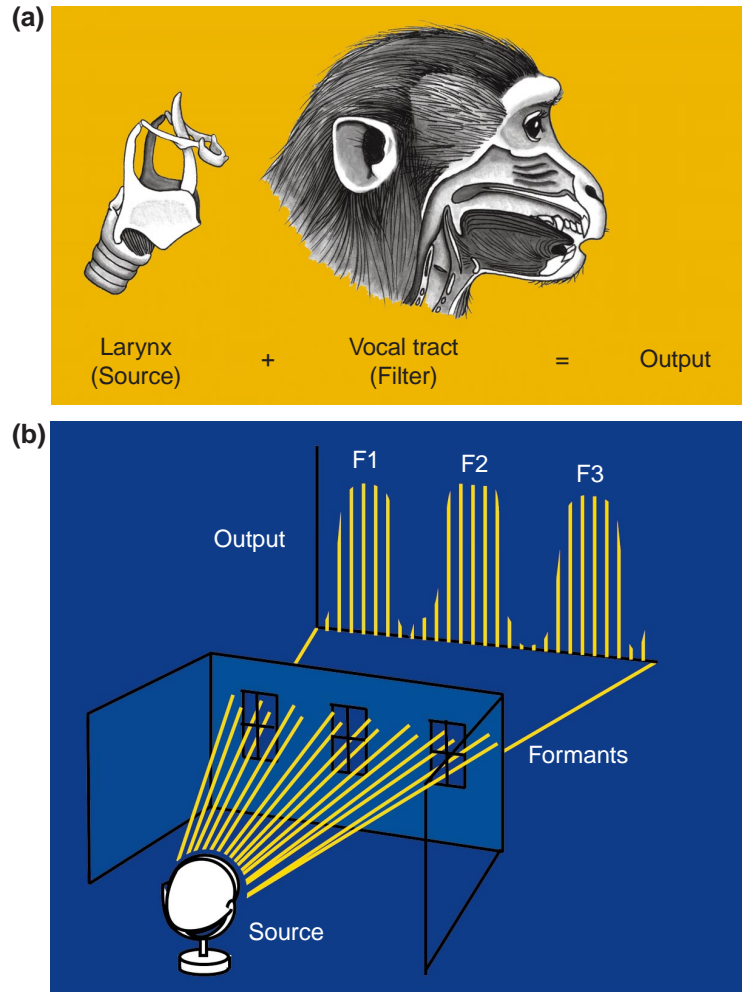
W.T. Fitch is at the Dept of Organismic and Evolutionary Biology, Harvard University, and in the Program in Speech and Hearing Science, Harvard University/MIT, Cambridge, MA 02138, USA.

tel: +1 617 496 6575  
fax: +1 617 496 8355  
e-mail: tec@wjh.harvard.edu

## Box 1. How vocal sounds are produced

Human speech uses rapid variations in various acoustic parameters to pack a startling amount of information into a short utterance. The basic machinery that underlies this process is very similar in humans and in other mammals: air exhaled from the lungs provides power to drive oscillations of the vocal folds (commonly known as vocal ‘cords’), which are located in the larynx or ‘voice box’. The rate of vocal fold oscillation (which varies from about 100 Hz in adult men to 500 Hz in small children) determines the pitch of the sound thus produced. The acoustic energy generated then passes through the vocal tract (the pharyngeal, oral and nasal cavities), where it is filtered, and finally out to the environment through the nostrils and lips. It is this filtering process that plays a crucial role in speech. The filtering is accomplished by a series of bandpass filters, which are termed formants. The formants modify the sound that is emitted, allowing specific frequencies to pass unhindered, but blocking the transmission of others. Formants are determined by the length and shape of the vocal tract, and are rapidly modified during speech by moving the articulators (tongue, lips, soft palate, etc.).

It is imperative to note that formants are independent of pitch. Pitch is determined by the vibration rate of the vocal folds (the source), whereas formants are determined by the vocal tract (the filter). The independence of source and filter is one of the key insights of modern speech acoustics, which is dubbed the ‘source/filter theory’ as a result (see Fig. 1). Vocal production is fundamentally different from most wind instruments (flutes, trumpets, clarinets, etc.), in which the pitch is determined by the resonances of the air column. This difference has been a source of pervasive confusion in both laymen and scientists. The situation is not helped by the fact that, although everyone knows what pitch is, we have no vernacular term for the perceptual correlate of formants. They are, broadly speaking, one



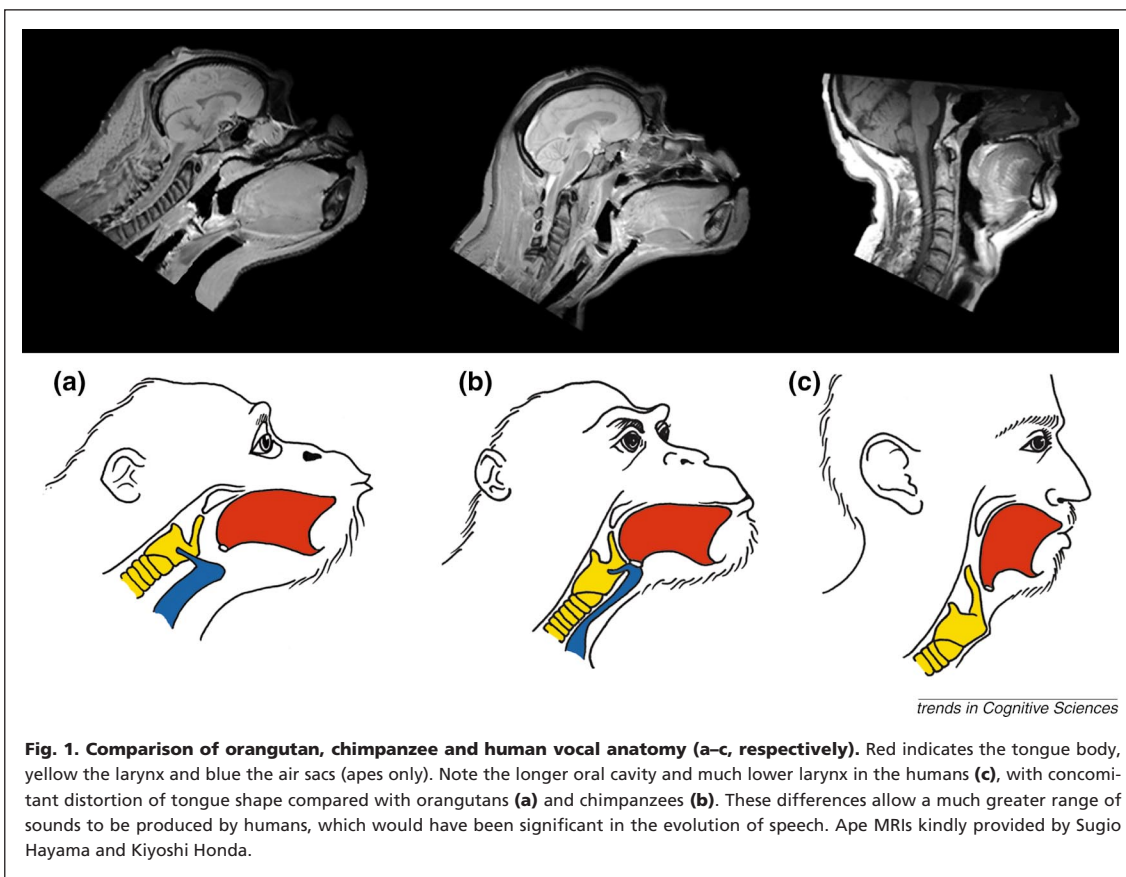
*trends in Cognitive Sciences*

**Fig. 1. Source/filter theory of vocal production.** The source/filter theory of vocal production, originally proposed for speech, appears to apply to vocal production in all mammals studied so far. The theory holds that vocalizations result from a sound source (typically produced at the larynx) combined with a vocal tract filter (which consists of a number of formants) (a). The formants, or vocal tract resonances, function as bandpass filters; they act as frequency ‘windows’ (b), allowing specific frequencies to pass through and blocking the transmission of others. This filtering action applies regardless of the type(s) of sound produced at the larynx.

component of ‘timbre’ or ‘voice quality’, but these terms are too broad to help much, and incorporate many acoustic parameters unrelated to formants. Nonetheless, formants are highly audible and salient. The difference between ‘beer’, ‘boot’, ‘bought’ and ‘bat’ is a difference in formants (primarily the lowest two formants, F1 and F2), and is clearly audible to both humans and other animals.

Several distinct abilities needed to converge before complex speech abilities became available to our ancestors. No single factor (or single mutation) could be solely responsible. At a minimum, speech production required changes in both peripheral mechanisms (relating to vocal acoustics and anatomy) and central neural mechanisms (those that underlie vocal control and imitation). With respect to the timing of these innovations, the fossil data have proven inconclusive,

and the traditional focus on such data has diverted attention from questions that can be addressed using data from living species. In particular, data from nonhuman primates allow us to infer ancestral functions of vocalization in early hominids, whereas data from more distantly related species that have convergently evolved vocal imitation (cetaceans, seals and birds) allow us to generate and test hypotheses about the function or functions of vocal learning.



**Fig. 1. Comparison of orangutan, chimpanzee and human vocal anatomy (a–c, respectively).** Red indicates the tongue body, yellow the larynx and blue the air sacs (apes only). Note the longer oral cavity and much lower larynx in the humans (c), with concomitant distortion of tongue shape compared with orangutans (a) and chimpanzees (b). These differences allow a much greater range of sounds to be produced by humans, which would have been significant in the evolution of speech. Ape MRIs kindly provided by Sugio Hayama and Kiyoshi Honda.

#### Peripheral differences: formants, speech and the descent of the larynx

Although one can argue that the evolution of language was independent of communicative mechanisms<sup>10</sup>, the evolution of speech was closely tied to mechanisms of sound production and perception<sup>9,11</sup>. Thus, the study of our species-typical communication system inevitably requires a basic knowledge of speech acoustics (see Box 1) and anatomy. This is especially true because the most obvious speech-related difference between humans and other mammals concerns the structure of the human vocal tract.

The study of the evolution of speech took off in the late 1960s, after major breakthroughs in the understanding of speech acoustics and perception<sup>12–14</sup>. A crucial first step was the recognition of the central importance of vocal tract resonances, or formants, in human speech. Formants function as bandpass filters, taking whatever sound emanates from the larynx, and shaping its spectrum into a series of peaks and valleys. All mammals that have been studied produce sounds in essentially the same way, using similar larynges and vocal anatomy, and all have vocal tracts with formants. However, humans make unusually heavy use of formants: they are the single most important acoustic parameter in human speech. This is clearly illustrated by whispered speech, in which the larynx generates broadband noise (with no vibration), but vocal tract movements are normal<sup>15</sup>. Whispered speech has no pitch, but is still clearly intelligible, because the formants are still present and normal. Another example is sinewave speech, a type of synthesized ‘speech’ that eliminates all acoustic cues except formant frequencies<sup>16</sup>. Although such signals sound like strange noises,

the linguistic message is nonetheless clearly intelligible to most people. The recognition of the central importance of formants paved the way for breakthroughs in speech science, and insights into the evolution of speech.

A central puzzle in the evolution of speech revolves around the fact that human vocal tract anatomy differs from other primates. Figure 1 shows midsagittal sections through the heads of an orangutan, a chimpanzee and a human obtained using MRI. It is evident that the human larynx rests much lower in the throat than in the apes. 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<sup>17</sup>. This is also the case in human infants, who can suckle (orally) and breathe (nasally) simultaneously. During human ontogeny, starting at about three months of age, the larynx begins a slow descent to its lower adult position, which it reaches after three to four years<sup>18</sup>. A second, smaller descent occurs in human males at puberty<sup>19,20</sup>. A similar ‘descent of the larynx’ must have occurred over the course of human evolution.

Nineteenth-century anatomists were aware of the uniqueness of the human vocal tract, but the acoustic significance of this configuration was not recognized until the 1960s, when speech scientist Lieberman and colleagues realized that the lowered larynx allows humans to produce a much wider range of formant patterns than other mammals<sup>21</sup>. 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. By varying the area of the oral and pharyngeal tubes independently, we can create a wide variety of vocal tract shapes

and formant patterns. By 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, and have an important role in allowing rapid, efficient speech communication to take place. Equally important, early workers in speech perception showed that speech relies on a unique encoding system that allows a much higher rate of data transmission than is possible with non-speech sounds<sup>22,23</sup>. Decoding this signal requires formant normalization, which is achieved most effectively using the vowel /i/ (Refs 11,24). Together this work established a causal connection between speech anatomy and phonetic ability: the low position of the adult human larynx enables us to produce sounds that have different, highly discriminable, formant patterns with ease. Thus, the descent of the larynx was a key innovation in the evolution of speech.

Although rarely noted, an equally striking aspect of human vocal anatomy is our lack of laryngeal air sacs. All great apes, and many other primates, have inflatable, soft-walled air pouches that extend out from the larynx and beneath the skin of the neck and thorax<sup>17,25</sup>. These sacs can hold up to 6 L of air and almost certainly serve a vocal function, but virtually nothing is known about their acoustic effects or adaptive significance. Perhaps they play a role in loud calls<sup>26,27</sup> but not in the type of quiet vocal interaction that typifies human communication<sup>28</sup>. Unfortunately, until more is known about the function of air sacs in living species, it is premature to speculate about their loss in our hominid ancestors. Nonetheless, the loss of air sacs in humans is as noteworthy as our gain of a descended larynx.

#### Neural differences: motor control and vocal imitation

In addition to a vocal tract that is anatomically capable of producing a large variety of formant patterns, human speech requires sophisticated nervous control. The most obvious aspect of this is the possibility that speech requires enhanced motor control over the vocal articulators (tongue, lips, velum, jaw, etc.). The fine, rapid motions of the tongue body that modify formant frequencies must be closely synchronized with other articulators (such as the lips and palate) as well as the vibrations of the larynx. The distinction between ‘pat’ and ‘bat’ is primarily a question of when the larynx begins vibrating relative to vocal tract movements, and a difference of tens of milliseconds is enough to differentiate these sounds perceptually. Thus, it is certainly plausible that human speech requires enhanced motor control (e.g. an increased motor neuron to vocal muscle fiber ratio). However, our understanding of animal vocal production is far less advanced than speech science, which makes it difficult to evaluate this hypothesis objectively. Despite major recent advances<sup>29–32</sup>, a quantitative comparison of the degree of vocal motor control in humans and animals is still unavailable.

Another important control issue concerns the hierarchical organization of speech segments (consonants and vowels) into higher-order structures (syllables, words and sentences). Such phonological structure lies at the border between speech *per se* and language, and is necessary to produce utterances of arbitrary complexity. The evolution of the ability to produce (and understand) such hierarchically organized streams of phonemes presumably presented a significant evo-

lutionary hurdle for our forebears. MacNeilage suggests that the evolutionary (and ontogenetic) precursor of syllabic structure was the mandibular oscillation associated with chewing and sucking, which provides a ‘frame’ onto which the ‘content’ of specific phonemes is superimposed<sup>33</sup>. Furthermore, on the basis of lesion data, it is proposed that different neural circuits underlie frame and content. Although the general framework MacNeilage offers is appealingly evolutionary and comparative, his neuroanatomical proposals have been vigorously debated, and this promises to be an active area of research.

However, there is one clear and undisputed difference between human vocal control and that of other primates. We are consummate vocal imitators, easily learning to produce whatever speech sounds we grow up with, together with musical sounds like singing and whistling. In sharp contrast, no nonhuman primates can learn to produce numerous sounds outside their ordinary species-specific repertoire<sup>34,35</sup>. Attempts to change the vocal repertoires of monkeys by cross-fostering them with other species have been disappointing<sup>36,37</sup>. Although evidence for vocal matching in primates exists<sup>38</sup>, and primates can be trained, with difficulty, to modify their calls<sup>39</sup>, the amount of acoustic variability observed is trivial compared with that necessary for human speech or song. Even chimps raised in human families, with extensive training and abundant rewards, fail to produce more than a few spoken words. By contrast, many studies have demonstrated that apes have the capacity to learn new gestures, pair them dependably with meanings and use them communicatively. Despite these good communicative abilities, and a capacity for perceptual learning of new sound–meaning pairings, the ability of nonhuman primates to produce learned sounds is limited or nonexistent.

This fact is made more curious by the abundant documentation of vocal imitation in nonprimate species<sup>35</sup>. Although evidence for vocal learning exists in aquatic mammals (seals and cetaceans), the superstars in this arena are clearly passerine birds. Avian song learning has been intensively studied since Marler’s groundbreaking work<sup>40,41</sup>, and recent research into the neural and genetic basis of song learning has been extremely fertile<sup>42</sup>. All oscines (‘true songbirds’) studied show some degree of song learning, requiring exposure to conspecific song early in life in order to develop normal songs themselves<sup>43</sup>. Specific groups, notably mimics like mockingbirds, take this much further, imitating the songs of other bird species, along with environmental sounds like crickets, creaking doors, car alarms and mobile phones. Finally, a variety of species is able to mimic human speech to a remarkable degree, and highly trained parrots have large vocabularies of speech sounds that they use communicatively<sup>44</sup>. Thus, when it comes to accomplished vocal imitation, humans are members of a strangely disjoint group that includes birds and aquatic mammals, but excludes our nearest relatives, the apes and other primates. Such vocal learning plays a crucial role in articulate speech, helping to generate the large vocabulary required by language<sup>45</sup>, and appears to represent a second key innovation in the evolution of spoken language<sup>28</sup>.

To summarize, we can isolate at least two indisputable changes that were associated with the evolution of speech.

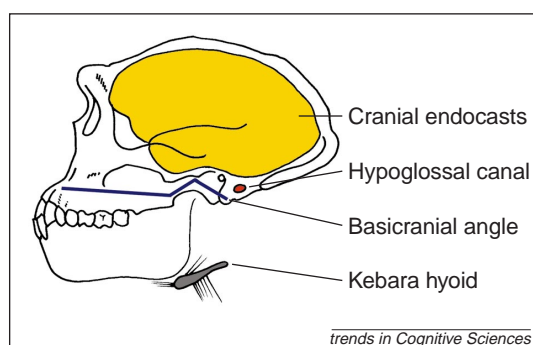
One is a change in the vocal production mechanism itself: the descent of the larynx enabled us to produce a variety of clearly discriminable formant patterns, and thus a vocal stream of adequate intricacy to communicate highly complex linguistic concepts. Second, the ability to imitate novel sounds vocally is a prerequisite for the formation of large vocabularies that typify all human languages. This ability is often taken for granted, but in fact is very unusual among mammals and requires an evolutionary explanation.

#### The fossil record: when did key innovations occur?

It is an unfortunate fact that speech does not fossilize, because it would clearly be useful to know when the innovations described above appeared. Physical anthropologists have attempted for many years to deduce when speech appeared by identifying fossil correlates of modern human vocal anatomy. Such an approach has been very successful in other domains: because the structure of the pelvis, leg and foot provided clear indicators of bipedality, fossil finds like 'Lucy' were revolutionary in showing that bipedalism preceded large brains by a million years or more. Unfortunately, no such straightforward links exist between skeletal morphology and vocal tract anatomy, because the vocal tract is a mobile structure that essentially floats in the throat, suspended from the skull by elastic ligaments and muscles. Thus, the morphology of the skull and hyoid bone (the only parts of vocal anatomy that fossilize) provide, at best, indirect clues about the capabilities of the vocal tract and the position of the larynx. Given the importance of knowing when key innovations occurred, however, physical anthropologists have generated a number of candidate examples of bony indicators of phonetic capabilities (Fig. 2).

The first possibility, recognized by Crelin and colleagues<sup>46–49</sup>, was that the flexion of the base of the skull reflects the position of the larynx, and thus that basicranial angle provides an indication of phonetic ability. Their measurements suggested that Neanderthals lacked a low larynx, and that, although they probably possessed some form of speech, their phonetic abilities were limited relative to anatomically modern *Homo sapiens* from the same time period. This suggestion, combined with evidence that Neanderthal tool culture was simple and relatively static<sup>50</sup>, suggested that Neanderthal language was less developed than ours, and that this might have played a role in their demise. Unfortunately, recent longitudinal X-ray data demonstrate no reliable correlation or causal connection between basicranial angle and larynx position in modern humans<sup>51</sup>. These data render any vocal tract reconstruction based solely on basicranial angle inconclusive.

The discovery of a well-preserved Neanderthal hyoid bone in Israel<sup>52</sup> also raises the question of Neanderthal speech abilities, because its anatomy is fundamentally modern. Its discoverers claimed that the Neanderthal vocal tract was also modern, with a descended larynx<sup>53,54</sup>. However, the morphology of the hyoid provides no reliable indication of the position of the larynx, and human hyoid anatomy shows no obvious changes as the larynx descends either in infancy, or later during puberty in males<sup>55</sup>. Although the Kebara hyoid thus appears mute as to the position of the Neanderthal larynx, its morphology suggests that Neanderthals did not



**Fig. 2. Proposed fossil indicators of extinct hominid vocal indicators.** Endocasts of the cranial interior have been used to estimate the size of gyri and sulci. The area of the hypoglossal canal has been used to estimate the size of the hypoglossal nerve which runs through it, providing innervation of the tongue muscle. Both basicranial angle (the angle of the line drawn through certain reference points on the skull base) and the morphology of the hyoid bone (from which the larynx is suspended; in this case the Kebara fossil hyoid, from a Neanderthal) have been used as an estimate of laryngeal height. None of these proposals appears to provide reliable indicators of the speech abilities of extinct hominids.

have air sacs, because the hyoid in chimps has a hollow into which the air sacs fit<sup>56</sup>.

A different approach involves using fossils to estimate the size of neural structures involved in spoken language. The first attempts, using brain endocasts to estimate the size of language-related cortical areas<sup>57,58</sup>, were of limited power, because surface features of the brain do not seem to provide clear indications of linguistic ability<sup>59</sup>. More recently however, Kay and colleagues used measurements of the size of the hypoglossal canal in modern primates and extinct hominids to draw conclusions about increased vocal control<sup>60</sup>. Because the hypoglossal nerve contains most of the motor fibers that innervate the tongue and other vocal articulators, these researchers made the reasonable assumption that a large hypoglossal canal would indicate a high ratio of motor neurons to tongue muscle fibers, and hence increased speech motor control. However, there is great variability in canal diameter between modern humans, with substantial overlap between measurements from humans and apes<sup>61</sup>. Thus, another proposed fossil diagnostic appears inadequate to deduce hominid speech capabilities with certainty.

In a recent attempt to tie fossil morphology to speech, MacLarnon and Hewitt analyzed measurements of the diameter of the thoracic vertebral canal in various extant primates, modern humans and extinct hominids<sup>62</sup>. They found that *Homo ergaster* ('*erectus*') and *Homo sapiens* have enlarged thoracic spinal cords, but that other primates and earlier hominids do not. Because the abdominal and intercostal muscles involved in breath control are supplied from thoracic motor neurons, the authors reasoned that the increased breath control associated with speech was present in these, but not in earlier, hominids. However, it is difficult to know whether increased respiratory control directly involved speech, or evolved for other reasons (e.g. prolonged running or swimming) and simply provided a necessary preadaptation to speech.

In summary, despite an extensive and disputatious literature, most potential fossil cues to phonetic abilities

appear inconclusive, suggesting that it will be difficult to reconstruct the vocal behavior of our extinct ancestors with any certainty. The quest for fossil evidence of language, and especially the question of Neanderthal speech, has been the dominant approach to studying the evolution of language for almost 30 years. This line of inquiry appears to have generated more heat than light, and diverted attention from alternative questions that are equally interesting and more accessible empirically. In particular, the question of when anatomical or neural changes took place is not the only one of interest; equally important is how and why they did. What we can say with certainty is that sometime in the approximately 6 million years since our divergence from chimps, major changes in hominid vocal anatomy and physiology took place, including a loss of air sacs and a descent of the adult larynx, together with acquisition of the ability to imitate novel sounds. What were the selective forces that led to these changes? How did they occur during our evolutionary history? These questions can be addressed by studying species that are still alive and vocalizing today.

### The comparative approach

In attempting to understand how and why human vocal communication diverged from that of other primates, it is imperative to adopt a comparative perspective. The comparative method provides a principled way to use empirical data from living animals to deduce the behavioral abilities of extinct common ancestors, together with clues to their adaptive function. Thus, study of the vocal behavior of non-human 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. Examples of convergent evolution (where similar traits have evolved independently in different lineages, presumably owing 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 currently weaker than is desirable. However, interest in non-human vocal production has increased dramatically recently, both in terms of peripheral or acoustic factors and central nervous control, and we can look forward to rapid advances in the near future. Recent comparative data bear on the origin and function of both peripheral and central adaptations that underlie human speech.

### Formants in animal communication and the descent of the larynx

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 non-human 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<sup>63,64</sup>. 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. Data from primates and birds demonstrate that animals can perceive formants with an accuracy rivaling that of humans<sup>65,66</sup>, and suggest that formants might play an important role in individual discrimination<sup>67</sup>. Thus, a role for formants as ‘vocal signatures’ might be widespread in vertebrate communication systems.

Formants might also provide an indication of the body size of a vocalizer. Vocal tract length is correlated positively with body size in humans, dogs and monkeys<sup>20,68,69</sup>. 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 different from voice pitch, which in fact has no correlation with body size in adult humans<sup>70</sup>. Together, these data suggest that our primate ancestors could have used formant frequencies to estimate body size from vocalizations. This in turn might have provided a preadaptation for ‘vocal tract normalization’, a crucial feature of speech perception whereby sounds from different-sized speakers are ‘normalized’ to yield equivalent percepts<sup>11,24,68</sup>.

The use of formants as cues to body size might have significance for the descent of the larynx as well. According to the widely accepted ‘phonetic expansion’ hypothesis, a low larynx permits a wider phonetic space. However, for the tongue to achieve the requisite freedom of movement, the larynx must be quite low, below the body of the tongue. What drove the presumably gradual descent of the larynx until it reached this point of phonetic advantage? Lieberman suggested that slight laryngeal lowering would be adaptive for mouth breathing during extreme physical challenge<sup>11</sup>, probably starting with *Homo erectus*. However, many mammals (e.g. dogs or cats) mouth breathe under stress, or for cooling by panting, without requiring any permanent larynx lowering. Another hypothesis was offered by DuBrul and others, who suggested that laryngeal lowering was a non-adaptive by-product of upright posture<sup>71</sup>. However, other habitually upright organisms, including arboreal species like gibbons and orangutans, or bipedal species like kangaroos, kangaroo rats or birds, show no lowering of the larynx. The ‘bipedal by-product’ hypothesis also fails to explain why, given the cost of a low larynx (vulnerability to choking), selection failed to correct this initially non-adaptive trait.

A different hypothesis is based on the fact that formants are correlated with body size<sup>68</sup>. One effect of a lowered larynx is to increase vocal tract length (and, consequently, to decrease formant frequencies). An animal with a lowered larynx can duplicate the vocalizations of a larger animal that lacks this feature, thus 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. Although Ohala initially offered this proposal as a refutation of Lieberman’s ‘phonetic expansion’ hypothesis<sup>72</sup>, the two are in fact compatible, with size exaggeration providing a pre-adaptation 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<sup>20</sup>. This second descent thus 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 and not specific to hominids, suggesting that other species might show vocal tract elongation as well. Comparative data suggest that other species have discovered a similar trick. Many bird species exhibit a peculiarity called tracheal elongation, in which 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 suggests that this serves to exaggerate the impression of size conveyed by vocalizations, which might be highly effective in animals that vocalize at night or from dense foliage<sup>73</sup>. A mammalian example of vocal tract elongation is provided by male red and fallow deer: during roar vocalizations they pull the larynx far down in the neck, sometimes as far as the thorax. This maneuver lowers the formants, and presumably increases the impressiveness of these roars, which serve to intimidate rivals and impress females during the mating season. Anatomical data suggest that descent of the larynx to exaggerate size might also be present in other mammals, such as lions<sup>74,75</sup>.

These comparative data suggest that a better understanding of animal communication and vocal production can lead to valuable insights into the evolution of speech. One important conclusion that follows from the size exaggeration hypothesis is that laryngeal lowering did not necessarily evolve in the context of improved speech production. Thus, even if fossil data could unambiguously pinpoint the time at which the larynx descended, the conclusion that this position indicates modern human speech capabilities would not follow. Further comparative work on animal vocal production and its communicative significance will allow us to test these and other hypotheses about the origins of the peripheral mechanisms that underlie speech. Of course, modern speech capabilities also required neural innovations.

### The function or functions of vocal imitation

The ability to listen to the vocal sounds of others and then imitate them is rare in mammals. However, such vocal learning is ubiquitous in songbirds, and in mammals it is found in humans, seals and cetaceans. What function or functions does vocal learning serve in those species that possess the ability?

One function of vocal learning in modern spoken language is obvious and crucial: to master language we must memorize a huge number of words that have essentially arbitrary sounds. All the disputants in the linguistic nature–nurture debate agree on this basic fact: vocal learning plays a crucial role in creating the extensive vocabulary upon which all spoken languages depend. However, this does not necessarily provide an explanation for the original function of vocal learning in our species, because a large vocabulary is a cultural artifact that probably requires vocal learning for its creation in the first place. Thus, we might expect that vocal learning originally proved adaptive in some other context, and then was co-opted for learning

large vocabularies once spoken language had already achieved some level of sophistication. For exploring such potential original functions of vocal mimicry, the comparative data set is very rich, mainly because of work with birds.

In nonhumans, the most obvious function of vocal learning is to create an elaborate vocal repertoire<sup>35</sup>. By accumulating and combining different songs (or song fragments) learned from multiple conspecifics, an individual can quickly develop a broad set of vocalizations that are within the species-typical range but also individually distinctive<sup>43</sup>. Such an elaborate repertoire could be useful for increasing attractiveness or territorial effectiveness. Females courted by a male that can continually produce different songs are less likely to habituate and move off. Consistent with this hypothesis, females of several passerine species prefer males with larger repertoires<sup>43</sup>. Alternatively, males with large repertoires might be more effective at defending territories for the same reason, or because continually varied song leads interlopers to conclude that more than one defender is present (the 'Beau Geste' hypothesis<sup>76</sup>). Either or both of these possibilities could also account for vocal learning in some whales and seals, in which males produce elaborate 'songs' while courting females and defending territories. According to this hypothesis, vocal learning originated to generate vocal complexity as an end in itself, rather than being a vehicle for communicating complex concepts. Such meaningless complexity could have provided a necessary preadaptation for the vocal communication of complex semantic structures.

A second hypothesis is that elaborate learned vocalizations function as an indicator of group membership. Among vocal learners, there are many examples of 'dialects' shared by a population, social group or extended family. One example is the 'signature whistles' seen in bottlenosed dolphins: specific frequency contours that appear to allow individual recognition by voice alone<sup>77</sup>. Many males adopt their mother's signature whistle before emigrating to distant waters, which could theoretically allow brothers who have never met to recognize each other as kin. Similar dialects that signal group membership or kinship are seen in killer whales<sup>78</sup>. These cetaceans, like chimpanzees and contemporary humans, live in social groups characterized by within-group cooperation and competition between groups. Such social systems put a premium on reliable indicators of group membership, vocal or otherwise. They also encourage an ability in newcomers who have emigrated into a group to learn the shared vocal indicators or group membership or 'passwords', and so could select for vocal learning<sup>79</sup>. Once basic learning capabilities have evolved, however, they would also allow interlopers to master the shibboleth of an invaded territory quickly. Such code-breaking would in turn select for increased discrimination on the part of group members or increasingly complex and hard-to-master 'passwords' (or both). If the value of being a group member is high, and vocal indicators of group membership are important, the stage would be set for a runaway selection leading to increased vocal learning and finer perceptual tuning.

This 'password hypothesis' for the origin of vocal learning is compatible with the observation that modern humans use accent to differentiate readily between individuals raised in their natal environment and newcomers (whose language

skills are otherwise adequate for communication). Thus, both the exquisitely developed imitative capacities of children, and the perceptual abilities of adults, go beyond what would be necessary for a very high level of vocal communication. The hypothesis makes a number of testable predictions (e.g. individuals should be more likely to aid a stranger that shares their dialect than other strangers), and is consistent with data from other species such as cetaceans. It provides a different perspective from which to view neurological phenomena like ‘foreign accent syndrome’, where individuals appear to lose their native accent while retaining otherwise normal speech<sup>80,81</sup>. It also suggests that a closer look at chimps might provide more subtle evidence that individuals can learn the ‘passwords’ of a new group, as recent data on chimp culture and learning suggests<sup>82</sup>.

An alternative possibility suggested by Donald is that vocal learning is just one example of a domain-general mimetic ability of modern humans<sup>83</sup>. We imitate gestures, facial expressions, dances, cooking and dressing styles, and so on, and Donald argues that the resultant homogeneity of tribal behavior played an important role in group cohesion in early humans. However, there are reasons to think that vocal learning might have preceded such generalized mimesis in phylogeny. The evolution of vocal imitation in the auditory domain is much easier than other forms of imitation (e.g. imitating facial gestures), because an individual can hear its own vocal output, and compare this with its memory of other individuals’ vocal output. Such an ‘acoustic mirror’ is intrinsic to auditory and vocal communication but absent in visual displays, and could account for the widespread occurrence of vocal imitation (and absence of general mimesis) in other taxa.

The comparative analysis above suggests that data from other species, even distantly related ones like birds or cetaceans, might help inform, clarify and test our understanding of vocal learning in our own species. However, comparative studies also highlight our ignorance about vocal learning in humans. Although major breakthroughs have been made in understanding the neural circuitry that underlies vocal learning in songbirds<sup>42</sup>, the neural mechanisms that underlie human vocal learning abilities remain largely unexplored, perhaps because it is not widely recognized that these abilities are unusual<sup>84,85</sup>. Are specific brain regions or connections dedicated to vocal mimicry in our species? Are there consistent individual differences in such abilities, and if so what is their neural basis? To what extent are vocal learning abilities dissociable from other skills necessary to master language? All of these questions can and should be addressed empirically using the existing tools and techniques of neuroscience and psychology.

### Conclusions

The evolution of speech is widely viewed as a prerequisite to rapid, flexible linguistic communication, and to the concomitant development of social living and culture that played such a crucial role in the recent evolutionary success of our species. Despite a long history of attempts to use fossils to deduce the timing of key events in the evolution of speech, the current fossil data are ambiguous and inconclusive. However, a recent surge of interest in animal vocal

### Outstanding questions

- All great apes have large air sacs attached to their larynges, and our most recent shared ancestors presumably also did. What are the acoustic and communicative functions of these air sacs in apes? Why did our hominid ancestors lose them?
- Current thinking holds that, even with a human brain in control, a chimp vocal tract could not produce certain crucial speech sounds. To what degree do limitations on nonhuman vocal production result from peripheral morphology versus neural control mechanisms?
- What were the crucial evolutionary innovations that allowed us to move from the highly canalized and limited vocal system of our primate ancestors to the flexible, open-ended production capabilities of modern humans? Vocal imitation, present in humans but not in other primates, appears to be one key innovation. Increased breath control, and freedom from stimulus driven control of vocalization, might be others.
- What are the neural mechanisms involved in vocal learning and imitation? One approach to this problem would use subtractive brain imaging techniques to examine a vocal imitation task. Are the brain areas involved the ones that have shown recent explosive expansion in our species (such as prefrontal cortex or the cerebellum)? Are homologous areas enlarged in other vocally imitating mammals such as seals or dolphins?

production and its role in the evolution of communication<sup>9,86</sup> has provided rich new empirical data to be incorporated into our thinking on the evolution of speech. This comparative approach is extremely promising, and has already provided important insights into the role of formants in nonhuman vocalizations. Finally, a crucial neural ability that has received too little attention is our unusual ability to vocally imitate sounds. Humans are clear outliers from other primates in this respect, though we share the ability with more distantly related vertebrates. The neural basis and adaptive significance of human vocal imitation should provide fertile ground for future investigations.

### Acknowledgements

The author thanks Sugio Hayama, Kiyoshi Honda and Eric Nicolas for help with Fig. 1, and gratefully acknowledges comments by Kim Beeman, Rodney Brooks, Daniel Dennett, Julia Fischer, Asif Ghazanfar, David Haig, Marc Hauser, Philip Lieberman, Will Lowe, Lynn Stein, Dan Weiss and Uri Wilensky.

### References

- 1 Ralls, K. *et al.* (1985) Vocalizations and vocal mimicry in captive harbor seals: *Phoca vitulina*. *Can. J. Zool.* 63, 1050–1056
- 2 Hayes, K.J. (1951) *The Ape in Our House*, Harper
- 3 Savage-Rumbaugh, E.S. *et al.* (1993) Language comprehension in ape and child. *Monogr. Soc. Res. Child Dev.* 58, 1–221
- 4 Hewes, G.W. (1973) Primate communication and the gestural origin of language. *Curr. Anthropol.* 14, 5–24
- 5 Falk, D. (1980) Language, handedness, and primate brains: did the Australopithecines sign? *Am. Anthropol.* 82, 72–78
- 6 Corballis, M. (1992) On the evolution of language and generativity. *Cognition* 44, 197–226
- 7 Jackendoff, R. (1999) Possible stages in the evolution of the language capacity. *Trends Cognit. Sci.* 3, 272–279
- 8 Ghazanfar, A.A. and Hauser, M.D. (1999) The neuroethology of primate vocal communication: substrates for the evolution of speech. *Trends Cognit. Sci.* 3, 377–384
- 9 Hauser, M.D. (1996) *The Evolution of Communication*, MIT Press
- 10 Chomsky, N. (1968) *Language and Mind*, Harcourt Brace
- 11 Lieberman, P. (1984) *The Biology and Evolution of Language*, Harvard University Press



- 12 Stevens, K.N. and House, A.S. (1955) Development of a quantitative description of vowel articulation. *J. Acoust. Soc. Am.* 27, 484–493
- 13 Chiba, T. and Kajiyama, M. (1958) *The Vowel: Its Nature and Structure*, Phonetic Society of Japan
- 14 Fant, G. (1960) *Acoustic Theory of Speech Production*, Mouton and Co.
- 15 Tartter, V.C. and Braun, D. (1994) Hearing smiles and frowns in normal and whisper registers. *J. Acoust. Soc. Am.* 96, 2101–2107
- 16 Remez, R.E. et al. (1981) Speech perception without traditional speech cues. *Science* 212, 947–950
- 17 Negus, V.E. (1949) *The Comparative Anatomy and Physiology of the Larynx*, Hafner
- 18 Sasaki, C.T. et al. (1977) Postnatal descent of the epiglottis in man. *Arch. Otolaryngol.* 103, 169–171
- 19 Fant, G. (1975) Non-uniform vowel normalization. *Speech Trans. Lab. Q. Prog. Stat. Rep.* 2–3, 1–19
- 20 Fitch, W.T. and Giedd, J. (1999) Morphology and development of the human vocal tract: a study using magnetic resonance imaging. *J. Acoust. Soc. Am.* 106, 1511–1522
- 21 Lieberman, P. et al. (1969) Vocal tract limitations on the vowel repertoires of rhesus monkeys and other nonhuman primates. *Science* 164, 1185–1187
- 22 Liberman, A.M. et al. (1967) Perception of the speech code. *Psychol. Rev.* 74, 431–461
- 23 Liberman, A.M. and Mattingly, I.G. (1985) The motor theory of speech perception revised. *Cognition* 21, 1–36
- 24 Nearey, T. (1978) *Phonetic Features for Vowels*, Indiana University Linguistics Club
- 25 Schön Ybarra, M. (1995) A comparative approach to the nonhuman primate vocal tract: implications for sound production. In *Frontiers in Primate Vocal Communication* (Zimmerman, E. and Newman J.D., eds), pp. 185–198, Plenum Press
- 26 Gautier, J.P. (1971) Etude morphologique et fonctionnelle des annexes extra-laryngées des cercopithecinae; liaison avec les cris d'espacement. *Biol. Gabonica* 7, 230–267
- 27 Schön Ybarra, M. (1988) Morphological adaptations for loud phonation in the vocal organ of howling monkeys. *Primate Rep.* 22, 19–24
- 28 Nottebohm, F. (1976) Vocal tract and brain: a search for evolutionary bottlenecks. *Ann. New York Acad. Sci.* 280, 643–649
- 29 Hauser, M.D. et al. (1993) The role of articulation in the production of rhesus monkey (*Macaca mulatta*) vocalizations. *Anim. Behav.* 45, 423–433
- 30 Hauser, M.D. and Schön Ybarra, M. (1994) The role of lip configuration in monkey vocalizations: experiments using xylocaine as a nerve block. *Brain Lang.* 46, 232–244
- 31 Jürgens, U. (1992) Neurobiology of vocal communication. In *Nonverbal Vocal Communication: Comparative and Developmental Approaches* (Papoušek, H. et al., eds), pp. 31–42, Cambridge University Press
- 32 Suthers, R.A. (1988) The production of echolocation signals by bats and birds. In *Animal Sonar: Processes and Performance* (Nachtigall, P.E. and Moore, P.W.B., eds), pp. 23–45, Plenum Press
- 33 MacNeilage, P.F. (1998) The frame/content theory of evolution of speech production. *Behav. Brain Sci.* 21, 499–546
- 34 Snowdon, C.T. (1990) Language capacities of nonhuman animals. *Yearbook of Physical Anthropology* 33, 215–243
- 35 Janik, V.M. and Slater, P.B. (1997) Vocal learning in mammals. *Adv. Stud. Behav.* 26, 59–99
- 36 Masataka, N. and Fujita, K. (1989) Vocal learning of Japanese and rhesus monkeys. *Behaviour* 109, 191–199
- 37 Owren, M.J. et al. (1993) Vocalizations of rhesus (*Macaca mulatta*) and Japanese (*M. fuscata*) macaques cross-fostered between species show evidence of only limited modification. *Dev. Psychobiol.* 26, 389–406
- 38 Elowson, A.M. and Snowdon, C.T. (1994) Pygmy marmosets, *Cebuella pygmaea*, modify vocal structure in response to changed social environment. *Anim. Behav.* 47, 1267–1277
- 39 Larson, C.R. et al. (1973) Sound spectral properties of conditioned vocalizations in monkeys. *Phonetica* 27, 100–112
- 40 Marler, P. (1970) A comparative approach to vocal learning: song development in white-crowned sparrows. *J. Comp. Physiol. Psychol.* 71, 1–25
- 41 Marler, P. (1991) Song learning behavior: the interface with neuroethology. *Trends Neurosci.* 14, 199–206
- 42 Nottebohm, F. (1999) The anatomy and timing of vocal learning in birds. In *The Design of Animal Communication* (Hauser, M.D. and Konishi, M., eds), pp. 63–110, MIT Press
- 43 Catchpole, C.K. and Slater, P.L.B. (1995) *Bird Song: Themes and Variations*, Cambridge University Press
- 44 Pepperberg, I.M. (1991) A communicative approach to animal cognition: a study of conceptual abilities of an African grey parrot. In *Cognitive Ethology* (Ristau, C.A., ed.), pp. 153–186, Lawrence Erlbaum
- 45 Nowak, M.A. and Krakauer, D.C. (1999) The evolution of language. *Proc. Natl. Acad. Sci. U. S. A.* 96, 8028–8033
- 46 Lieberman, P. and Crelin, E.S. (1971) On the speech of Neanderthal man. *Linguist. Inq.* 2, 203–222
- 47 Lieberman, P. et al. (1972) Phonetic ability and related anatomy of the newborn and adult human, Neanderthal man, and the chimpanzee. *Am. Anthropol.* 74, 287–307
- 48 Laitman, J.T. and Crelin, E.S. (1976) Postnatal development of the basicranium and vocal tract region in man. In *Symposium on Development of the Basicranium* (Bosma, J.F., ed.), US Govt Printing Office
- 49 Crelin, E. (1987) *The Human Vocal Tract*, Vantage Press
- 50 Mellars, P. (1998) Neanderthals, modern humans and the archaeological evidence for language. In *The Origin and Diversification of Language* (Jablonski, N.G. and Aiello, L.C., eds), California Academy of Sciences
- 51 Lieberman, D.E. and McCarthy, R.C. (1999) The ontogeny of cranial base angulation in humans and chimpanzees and its implications for reconstructing pharyngeal dimensions. *J. Hum. Evol.* 36, 487–517
- 52 Arensburg, B. et al. (1989) A middle paleolithic human hyoid bone. *Nature* 338, 758–760
- 53 Arensburg, B. et al. (1990) A reappraisal of the anatomical basis for speech in middle Paleolithic hominids. *Am. J. Phys. Anthropol.* 83, 137–146
- 54 Arensburg, B. (1994) Middle Paleolithic speech capabilities: a response to Dr. Lieberman. *Am. J. Phys. Anthropol.* 94, 279–280
- 55 Lieberman, P. et al. (1989) Folk physiology and talking hyoids. *Nature* 342, 486
- 56 Schultz, A.H. (1969) The skeleton of the chimpanzee. In *The Chimpanzee* (Vol. 1) (Bourne, G., ed.), pp. 165–187, Karger
- 57 LeMay, M. (1975) The language capability of Neanderthal man. *Am. J. Phys. Anthropol.* 42, 9–14
- 58 Falk, D. (1980) Language, handedness, and primate brains: did the Australopithecines sign? *Am. Anthropol.* 82, 72–78
- 59 Gannon, P.J. et al. (1998) Asymmetry of chimpanzee planum temporale: humanlike pattern of Wernicke's brain language area homolog. *Science* 279, 220–222
- 60 Kay, R.F. et al. (1998) The hypoglossal canal and the origin of human vocal behavior. *Proc. Natl. Acad. Sci. U. S. A.* 95, 5417–5419
- 61 DeGusta, D. et al. (1999) Hypoglossal canal size and hominid speech. *Proc. Natl. Acad. Sci. U. S. A.* 96, 1800–1804
- 62 MacLarnon, A. and Hewitt, G. (1999) The evolution of human speech: the role of enhanced breathing control. *Am. J. Phys. Anthropol.* 109, 341–363
- 63 Hauser, M.D. (1992) Articulatory and social factors influence the acoustic structure of rhesus monkey vocalizations: a learned mode of production? *J. Acoust. Soc. Am.* 91, 2175–2179
- 64 Rendall, D. et al. (1996) Vocal recognition of individuals and kin in free-ranging rhesus monkeys. *Anim. Behav.* 51, 1007–1015
- 65 Owren, M.J. (1990) Acoustic classification of alarm calls by vervet monkeys (*Cercopithecus aethiops*) and humans: II. Synthetic calls. *J. Comp. Psychol.* 104, 29–40
- 66 Sommers, M.S. et al. (1992) Formant frequency discrimination by Japanese macaques (*Macaca fuscata*). *J. Acoust. Soc. Am.* 91, 3499–3510
- 67 Fitch, W.T. and Kelley, J.P. (2000) Perception of vocal tract resonances by whooping cranes, *Grus americana*. *Ethology* 106, 559–574
- 68 Fitch, W.T. (1997) Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *J. Acoust. Soc. Am.* 102, 1213–1222
- 69 Riede, T. and Fitch, W.T. (1999) Vocal tract length and acoustics of vocalization in the domestic dog, *Canis familiaris*. *J. Exp. Biol.* 202, 2859–2869
- 70 Künzel, H.J. (1989) How well does average fundamental frequency correlate with speaker height and weight? *Phonetica* 46, 117–125

- 71 DuBrul, E.L. (1976) Biomechanics of speech sounds. *Ann. New York Acad. Sci.* 280, 631–642
- 72 Ohala, J.J. (1984) An ethological perspective on common cross-language utilization of F<sub>0</sub> of voice. *Phonetica* 41, 1–16
- 73 Fitch, W.T. (1999) Acoustic exaggeration of size in birds by tracheal elongation: comparative and theoretical analyses. *J. Zool.* 248, 31–49
- 74 Pocock, R.I. (1916) On the hyoid apparatus of the lion (*F. leo*) and related species of Felidae. *Ann. Mag. Nat. Hist.* 8, 222–229
- 75 Hast, M. (1989) The larynx of roaring and non-roaring cats. *J. Anat.* 163, 117–121
- 76 Krebs, J.R. (1977) The significance of song repertoires: the Beau Geste hypothesis. *Anim. Behav.* 25, 475–478
- 77 Sayigh, L.S. et al. (1990) Signature whistles of free-ranging bottlenose dolphins, *Tursiops truncatus*: stability and mother-offspring comparisons. *Behav. Ecol. Sociobiol.* 26, 247–260
- 78 Ford, J.K.B. (1991) Vocal traditions among resident killer whales (*Orcinus orca*) in coastal waters of British Columbia. *Can. J. Zool.* 69, 1454–1483
- 79 Feeke, F. (1982) Song mimesis within colonies of *Cacicus c. cela* (Icteridae: Aves): a colonial password? *Zeitschrift Tierpsychologie* 58, 119–152
- 80 Monrad-Krohn, G.H. (1947) Dysprosody or altered 'melody of language'. *Brain* 70, 405–415
- 81 Blumstein, S.E. et al. (1987) On the nature of the foreign accent syndrome: a case study. *Brain Lang.* 31, 215–244
- 82 Marshall, A.J. et al. (1999) Does learning affect the structure of vocalizations in chimpanzees? *Anim. Behav.* 58, 825–830
- 83 Donald, M. (1993) *Origins of the Modern Mind*, Harvard University Press
- 84 Deacon, T.W. (1997) *The Symbolic Species: The Co-evolution of Language and the Brain*, W.W. Norton
- 85 Tomasello, M. et al. (1993) Imitative learning of actions on objects by children, chimpanzees, and enculturated chimpanzees. *Child Dev.* 64, 1688–1706
- 86 Bradbury, J.W. and Vehrencamp, S.L. (1998) *Principles of Animal Communication*, Sinauer

# Social perception from visual cues: role of the STS region

Truett Allison, Aina Puce and Gregory McCarthy

**Social perception refers to initial stages in the processing of information that culminates in the accurate analysis of the dispositions and intentions of other individuals. Single-cell recordings in monkeys, and neurophysiological and neuroimaging studies in humans, reveal that cerebral cortex in and near the superior temporal sulcus (STS) region is an important component of this perceptual system. In monkeys and humans, the STS region is activated by movements of the eyes, mouth, hands and body, suggesting that it is involved in analysis of biological motion. However, it is also activated by static images of the face and body, suggesting that it is sensitive to implied motion and more generally to stimuli that signal the actions of another individual. Subsequent analysis of socially relevant stimuli is carried out in the amygdala and orbitofrontal cortex, which supports a three-structure model proposed by Brothers. The homology of human and monkey areas involved in social perception, and the functional interrelationships between the STS region and the ventral face area, are unresolved issues.**

Consider the predicament of Barbara Ehrenreich, who is considering a vacation out West<sup>1</sup>:

It would be nice to go on a vacation where I didn't have to worry about being ripped limb from limb by some big ursine slob...All right, I know the ecologically correct line: 'They won't bother you if you don't bother them.' But who knows what bothers a bear?...So instead of communing with the majestic peaks and flower-studded meadows, I spend my hikes going over all the helpful tips for

surviving an Encounter. Look them in the eye? No, that was mountain lions. Bears just hate it when you stare at them, so keep your gaze fixed dreamily on the scenery. Play dead? Let's see, that works for grizzlies but not for black bears. So do you take off the backpack, get out the wildlife guidebook, do a quick taxonomic determination and *then* play dead?

If it is difficult to infer the intentions of other humans from their facial gestures and body language, it is even harder,

T. Allison is at the Neuropsychology Laboratory, VA Medical Center, West Haven, CT 06516 and the Department of Neurology, Yale University School of Medicine, New Haven, CT 06510, USA.

tel: +1 203 932 5711  
fax: +1 203 937 3474  
e-mail: truett.allison@yale.edu

A. Puce is at the Brain Sciences Institute, Swinburne University of Technology, PO Box 218, Hawthorn, Victoria 3122, Australia.

e-mail: puce@bsi.swin.edu.au

G. McCarthy is at the Brain Imaging and Analysis Center, Box 3808, Duke University Medical Center, Durham, NC 27710, USA.

e-mail: gregory.mccarthy@duke.edu