Evolution of the human tongue and emergence of speech biomechanics

Axel G. Ekström* and Jens Edlund

Division of Speech, Music and Hearing, KTH Royal Institute of Technology, Stockholm, Sweden

The tongue is one of the organs most central to human speech. Here, the evolution and species-unique properties of the human tongue is traced, via reference to the apparent articulatory behavior of extant non-human great apes, and fossil findings from early hominids – from a point of view of articulatory phonetics, the science of human speech production. Increased lingual flexibility provided the possibility of mapping of articulatory targets, possibly via exaptation of manual-gestural mapping capacities evident in extant great apes. The emergence of the human-specific tongue, its properties, and morphology were crucial to the evolution of human articulate speech.

1. Introduction

The shape, proportions, and positioning of the human tongue are crucial components of speech biomechanics, accounting for both articulatory and temporal properties of speech acoustics and production. Here, something of the evolution of the organ and its properties is traced via reference comparative anatomy and speech-like behavior in extant non-human great apes (hereafter great apes), and archeological findings of extinct hominid morphology. A cohesive account of speech evolution must take account of both anatomical and neurological properties. In particular, models should seek to include and consider the evolution and properties of the human tongue.

2. Properties and morphology

The mammalian tongue is widely considered a muscular hydrostat (Smith and Kier, 1989; Gilbert et al., 2007; Takemoto, 2008), consisting of muscles with no skeletal support and performing hydraulic movements characterized by the property of being largely incompressible from physiological pressures. Functionally, the volume of a muscular hydrostat is constant, and compression in any dimension causes appropriate compensatory expansion in another. Across species, such properties facilitate mastication and swallowing, but in humans, deformation is a crucial component of speech production mechanics also. For speech evolution, thus, vocal anatomy may provide scholars with clues as to the nature of morphological changes that took place throughout human evolution before speech emerged.

The human tongue possesses four major extrinsic and four intrinsic muscles. The extrinsic muscles (originating outside of the organ itself) are: the genioglossus, responsible for forward

and downward movement of the tongue (anterior) and forward movement of the dorsal tongue body extending into the pharynx (posterior); the styloglossi, which retract the tongue; the hyoglossus, which depresses and retracts the tongue; and the palatoglossus, which elevates the posterior position of the tongue; and the four intrinsic muscles (attaching only to other muscles in the tongue body) are the superior longitudinal and inferior longitudinal and transverse and vertical muscles. This gross musculature is largely conserved across primates (Swindler and Wood, 1982; Takemoto, 2008) but the human tongue and face contain a higher proportion of slow-twitch myosin fibers, compared to other primates (Sanders et al., 2013; Burrows et al., 2014).

Compared to other mammals, in (adult) humans, the larynx – and therefore also the tongue root, as the larynx is suspended from the basihyoid bone – is permanently retracted downward into the throat (Negus, 1949; Lieberman, 1984, 2012; de Boer and Fitch, 2010; Lieberman et al., 2001). In comparison – as was noted by both Negus (1949) and Crelin (1987) – the tongues of nonhuman mammals, such as sheep, dogs, cats, macaques, spider monkeys, chimpanzees (as well as human infants, who achieve the adult configuration in childhood), are located entirely within the oral cavity. Studies of dissected specimens have since been complemented with studies of live vocalizing animals (Fitch 1997, 2000a; Fitch and Reby, 2001; Weissengruber et al., 2002) illustrating that nonhuman animals typically do not employ their tongues in vocalization. Further, while larynx lowering is found in other species (Fitch and Reby, 2001), these reflect wholly separate adaptations from that of humans. For example, the lowering of the larynx in the Red Deer studied by Fitch and Reby does not – and cannot – markedly change the corresponding phonetic range of the animal.

This is so because, in the words of Lieberman (2012, p. 612): "the larynx transiently descends [in deer] by increasing the distance between the hyoid bone and larynx. This maneuver does not change the shape of the SVT—its cross-sectional area function as a function of distance." The tongue remains firmly anchored in the mouth of the animal. In comparison, the descended position of the human larynx is part of a suite of extensive anatomical changes in evolution, involving the tongue's reshaping and partial descent into the pharynx, expansion of the pharyngeal cavity, and restructuring of the cranium (Negus, 1949; Bosma, 1975; Laitman et al., 1978; Laitman and Heimbuch, 1982; Lieberman and McCarthy, 1999). Thus, while Fitch (2000b) claims that a lowered larynx evolved to shift down resonance frequencies and provide impressions of greater size, this claim as applied to humans is based on a false equivalency. The claim is likely true of various nonhuman animals, including Red Deer (Fitch and Reby, 2001); however, the same mechanism does not explain larynx lowering in human vocal tracts. The two are functionally nonequivalent.

Through its reconfiguration, the human supralaryngeal vocal tract acquires a roughly 1:1 relationship between horizontal and vertical sections. The human tongue has been rounded, compared to that of the "flatter" tongues of nonhuman mammals (Negus, 1949; Crelin, 1987; Takemoto, 2008; Iwasaki et al., 2019). The resulting flexibility of tongue motion makes possible the production of *quantal* vowels including [i] and [u] (the vowels in "see" and "boot," respectively), and velar plosives [k] and [g] (the first consonants in "cup" and "good," respectively) (Lieberman et al., 1969, 1992; Stevens, 1989; Carre et al., 1995; de Boer and Fitch,

2010; Lieberman, 2012). In comparison, ascribing hydrostatic properties to the chimpanzee tongue indicates that its freedom of motion is primarily in protrusion and retrusion, as opposed to deformation dorsally inside the oral cavity (required for a variety of speech sounds) (Takemoto, 2008). Crucially, anterior degrees of freedom are necessary for achieving the full extent of human articulatory space (see, e.g., Engwall, 2003). For example, both [i] and [u] are high vowels, articulated with the tongue tip or body arched toward the palate, respectively (Figure 1). Thus, it is the relative position and shape the tongue, rather than position of the larynx *per se*, which are central for speech (Lieberman, 1984, 2012; Carre et al., 1995; de Boer and Fitch, 2010). No nonhuman mammal have ever been shown to attain the same configuration necessary for the extremities of human speech (Lieberman, 1984, 2012; de Boer and Fitch, 2010; Fitch et al., 2016; Ekström, 2023a).
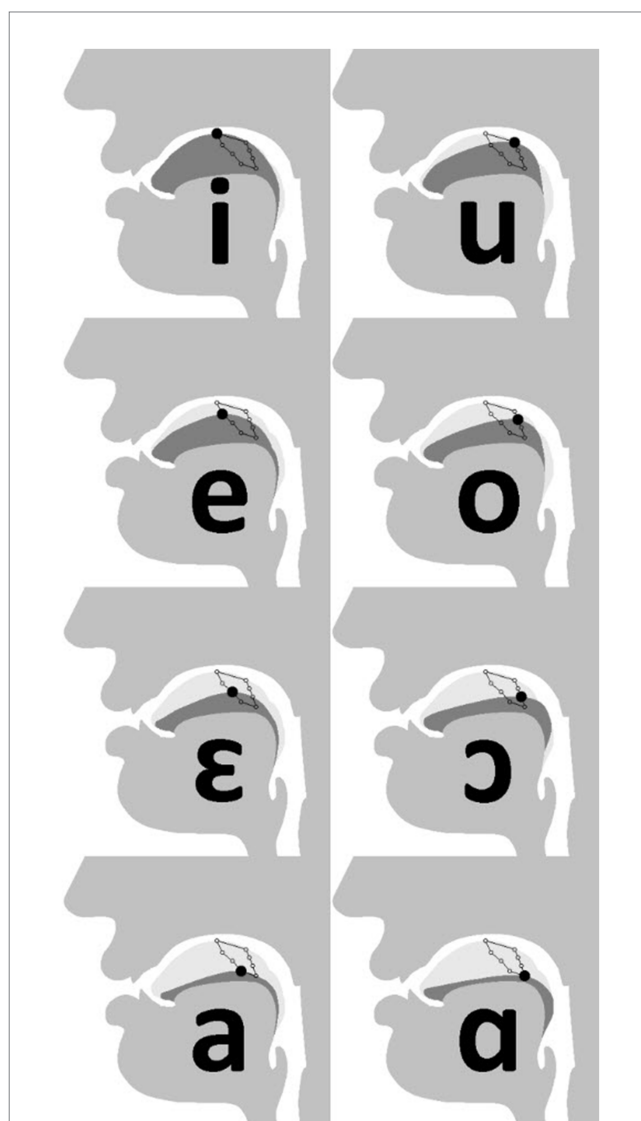

FIGURE 1
Tongue position for cardinal vowels. [i] (top-left), [u] (top-right) and [a] (bottom-left) are quantal vowels, produced in comparatively stable articulatory space (Lieberman, 1984, 2012; Stevens, 1989). Image adapted from Ishwar (2020). CC BY 3.0.

3. Evolutionary history

3.1. The primate tongue is deemphasized for food intake

While estimates differ, the lineage leading to modern humans is assumed to have diverged from that of *Pan* around ~7 Mya (White et al., 2009). The phylogenetically more distant Strepsirrhini (e.g., lemurs) possess lingual features markedly different from those of extant haplorrhines (Old World monkeys and apes, and New World monkeys) (Machida et al., 1967; Iwasaki et al., 2019). Strepsirrhine tongues possess a narrow lingual apex, anterior–posterior elongated outline, and developed sublingual (Fleagle, 2013), giving those species significant degrees of freedom outside the oral cavity for manipulation of food stuff. On the other hand, lingual anatomy of haplorrhines indicates a diminished role of the tongue in food uptake specifically, possibly coinciding with the emergence of opposable thumbs used for active manual manipulation of food (Ankel-Simons, 2007; Fleagle, 2013; Iwasaki et al., 2019).

Concurrently, the evolutionary trend of the hominid head, coinciding in phylogenetic history with a dietary shift from raw to processed and/or cooked foods (Wrangham, 2009) shows the emergence of species that spent less time masticating and gestating its food. Prognathia was reduced (the face pulled back toward the cranium), and dentition and the oral cavity were reduced in size, and the tongue reshaped. The position of the hyoid bone, providing the tongue with an osseous base, also shifted in evolution (Lieberman, 2011). For speech, the same sequence of changes seemingly "freed up" the facial muscles, organs, and larger would-be articulatory complex, scaffolding more extensive motor sequence cycles – i.e., complex syllabic speech (MacNeilage, 1998). The homo lineage represents an extreme of the haplorrhine trend, with manipulation and processing of food via tool use facilitating a near-complete outsourcing of food intake processes to the hands (Osvath and Gärdenfors, 2005; Wollstonecroft, 2011; Iwasaki et al., 2019).

3.2. Reconstructing speech capacities of extinct hominids

Seminal work on speech capacities of extinct hominids were performed by Lieberman and Crelin (1971) and Lieberman et al. (1972), who developed reconstructions of Neanderthal (*H. neanderthalensis*) vocal tracts. The authors reconstructed the supralaryngeal vocal tract of the La Chapelle-aus-Saints Neanderthal fossil, and simulated by means of a computer program, all possible vocal tract configurations. They found that the resulting vowel space was (1) greater than that estimated for actual chimpanzee vocalizations (which the authors attributed to the chimpanzee possibly lacking crucial neural mechanisms for fully utilizing the phonetic potential of species' vocal tracts); and (2) like the vowel space of human infants, it did not include quantal vowels [a], [i], or [u], which require extreme 10:1 midpoint discontinuities in the oral tract (Stevens, 1989; Lieberman, 2012).

Throughout the history of research on evolution of human speech-centric anatomy, no series of efforts are more extensive than those of Crelin (1987, 1989); throughout this undertaking, he became "convinced that our development is a résumé of our evolution" (Crelin,

1989, p. 19). Crelin determined that skulls of both australopithecines and *Homo habilis* were essentially "apelike" (Crelin, 1987). Further, based on comparative analysis including the skull of the Taung child – a juvenile *A. africanus* (Dart, 1925) – Crelin also argued that the vocal anatomical ontogenetic development of the genus was also essentially comparable to that of extant apes (see Nishimura, 2005).

Vocal tracts of *H. erectus* were deemed intermediate in form between the apelike vocal tracts of australopithecines, and that of the "modern" human vocal tracts: "The snout, related to relatively large maxillae, coupled with a relatively short robust mandibular ramus indicates that only a part of the posterior third of the tongue was located low enough in the neck to serve as a short anterior wall to the oropharynx" (Crelin, 1987, p. 158). Finally, following a restoration of the "Steinheim skull" of an adult archaic (likely female) human (*H. heidelbergensis*) estimated to around 250–350 Kya, Crelin noted that the fossil skull base was "identical to that of a present-day *Homo sapiens* skull." He determined that the archaic human represented by the Steinheim skull would have been capable of the full range of human speech sounds. Crelin's conclusions, then, was that the full extent of modern human speech capacities likely had not evolved until the emergence of archaic modern humans, but that something of the capacity had evolved with *H. erectus*.

Both the original reconstructions by Lieberman et al. (1972) and later works by Crelin (1987) assumed that flexion of the skull base (cranial-base angle) provided a basis for inferring the likely shape of species' vocal tracts: "A fossil that had a shallow cranial base similar to that seen in living apes and human newborns presumably had a similar vocal tract, while a fossil having a flexed adult human basicranial angle would have had a human vocal tract" (Lieberman, 2007a, p. 45). Such measurements are problematic, however, as the tongue and larynx continue to descend in humans after the ontogenetic point of stabilization of cranial flexure (Fitch and Giedd, 1999; Lieberman and McCarthy, 1999). It is important to note that this evidence had yet to be presented at the time of either the original reconstructions by Lieberman et al. (1972), or the later efforts by Crelin (1987) (see, e.g., Lieberman, 2007a). Nonetheless, it is of note that the central claims made on the basis of those studies – that Neanderthal speech was likely less articulate than that of modern humans, resulting from its not yet having acquired the supralaryngeal airway dimensions that characterize the human condition – are also seemingly supported by other findings, including the observation that fitting a human vocal tract (with a 1:1 relationship between horizontal and vertical sections) to Neanderthal anatomy effectively places the larynx in the chest, a vocal anatomical configuration absent from any existent mammal (see Lieberman, 2007b).

3.2.1. Alternate views

The most widely discussed purported refutation of the findings of Lieberman et al. (1972) is that of Boë et al. (1999), contextually an important work, as it is the only one to couch is suppositions in speech production and acoustics (*cf.* Carlisle and Siegel, 1974; D'Anastasio et al., 2013; Dediu and Levinson, 2013; Frayer, 2017). However, like the earlier reconstructions by Lieberman et al., Boë et al. based their research and argument – that "Neanderthal man was not morphologically handicapped for speech" – on angle of the cranial base (in their work, of a reconstruction of a Neanderthal skull by Heim, 1989). In so doing, however, the authors, while citing the then-recent findings that tongue position and shape could not be inferred

from basicranial angle (Fitch and Giedd, 1999; Lieberman and McCarthy, 1999), fail to acknowledge their importance (see Lieberman, 2007a). Boë et al. also fit a human vocal tract to the reconstructed Neanderthal skull. In commenting on this procedure, Lieberman (2007b, p. 552) writes, "The restructuring of the human skull which places the human face in line with the braincase did not take place in Neanderthals, resulting in a long oral cavity. A modern vocal tract placed on a Neanderthal skull would require a tongue displaced down so low into its neck that the creature's larynx would be in its chest, a configuration absent in any primate species."

A second source of error in the Boë series of works relates to the "Variable Linear Articulatory Model" (VLAM) procedure, based on an algorithm by Maeda (1990), and consistently employed by Boë et al. throughout their work on the topic. By the logic of this algorithm, research teams led by Boë argued there were no anatomical limitations to Neanderthals' (Boë et al., 1999, 2002a) or human infants' (Boë et al., 2002b, 2007) producing the full range of human speech, and that the size of the pharynx was "an irrelevant parameter for speech emergence" (Boë et al., 2002b). Crucially, however, the Maeda algorithm – based on adult human French speakers – maintains the basic shape of the human supralaryngeal vocal tract, even if it results in anatomically impossible vocal tract configurations – as was indeed the case. The problematic application of the Maeda algorithm was outlined by de Boer and Fitch (2010), see also Lieberman (2007a, 2012). After these refutations, no further work using the VLAM procedure have been produced by teams led by Boë.

The reconstructions by Lieberman and Crelin have been criticized at various times and by various researchers. However, much of the debate have not focused on elements of speech production *per se*, but rather on a deeper anthropological consideration of whether Neanderthal should be considered a separate species from *homo sapiens* (*cf.* a subspecies; *Homo sapiens neanderthalensis*). This debate is largely outside the scope of this text and will not be discussed beyond this point.[1] However, it is important to note that various authors of critiques of the speech-centric work by Lieberman et al. fail to address basic tenets of the relevant anatomical arguments. For example, Dediu and Levinson (2013), quoting Fitch (2009, p. 133) cite the dynamic lowering of the larynx in nonhuman animals as evidence that the "significance of the descent of the larynx … has been overestimated." It has already been shown (in section "Properties and morphology"), why this is a non-argument: larynx lowering (permanent or temporary) in nonhuman animals is not functionally equivalent to that found in humans. The two are accomplished disparately, and for different purposes (Fitch and Reby, 2001; Lieberman, 2012), and ontogenetic laryngeal descent in humans is part of a suite of anatomical changes facilitating speech capacities, including a restructuring of the cranium and expansion of the pharyngeal cavity. Dediu and Levinson (2013) also uncritically cite the widely discredited modeling work by Boë et al. (1999) as positive proof against the Lieberman claims (for refutations, see de Boer and Fitch, 2010; Lieberman, 2012).

3.2.2. Speaking hyoids?

The hyoid bone constitutes one of the least represented elements in the fossil record, with the only known findings representing *Australopithecus aferensis* (Alemseged et al., 2006), *H. heidelbergensis* (Martínez et al., 2008), and *H. neanderthalensis* (Neanderthals) (Arensburg et al., 1989). Nevertheless, findings provide useful clues to the evolution of the modern human articulatory apparatus. The human hyoid is bar-shaped and positioned below the tongue, under the inferior margin of the mandibular body, while that of extant great apes is bulla-shaped, and positioned anterior to the tongue root (Falk, 1975; Steele et al., 2013). Comparative studies suggest that the hyoid of *A. afarensis* is bulla-shaped like that of extant great apes (Alemseged et al., 2006) (though in the specimen described, the hyoid was preserved beneath the palate, preventing thorough analysis). Meanwhile, general morphology of the *H. heidelbergensis* hyoid bones found at Sima de los Heuses and described by Martínez et al. (2008) (dated to ~530 Kya), shows a transition away from the bulla-shaped hyoid of extant nonhuman hominids, toward the bar-shaped morphology that characterizes the hyoid of modern humans. The authors suggested that such aspects of modern hyoid bone morphology are a derived feature, inherited from a common ancestor of the Neanderthal and modern humans (Arensburg et al., 1989; Martínez et al., 2008). This determination is consistent with Crelin's (1987) constructions. Crucially for all such work, however, the shape of the hyoid is maintained in human infants and adults, even as the hyoid and larynx descend in ontogenetic development. Thus, as was argued by Lieberman (1999), the shape of the hyoid *per se* does not, and cannot, inform researchers about the length or shape of the vocal tract of extinct hominids (see also Lieberman et al., 1989), and therefore provides only circumstantial evidence with bearing on actual articulation.[2]

3.2.3. Hypoglossal canals

It has been suggested, based on studies of the hypoglossal canal of the occipital bone (cranial nerve XII), which transmits the nerve supplying all intrinsic and (all but one) extrinsic lingual muscles, that mean areas of the hypoglossal canal of humans is significantly larger than that of other extant hominids (Kay et al., 1998). However, DeGusta et al. (1999) showed that hypoglossal canal size was highly variable in humans, with overlap between modern humans, and both extant nonhuman great apes, and australopithecines. The same conclusions were later enforced by Jungers et al. (2003), see also Lieberman (1999). Thus, the current state of research does not support that the shape of hypoglossal canals provides reliable information about species' speech capacities.

3.3. Summary

To date, it has never been convincingly argued that any other species than *homo sapiens* possessed the full range of modern human speech capacities. While individual elements of anatomy do not in isolation provide researchers with the necessary information for

---

1 For example, the arguments by Carlisle and Siegel (1974) are refuted by Lieberman and Crelin (1974); claims by Dediu and Levinson (2013) are challenged by Berwick et al. (2013).

2 Further, little work to date has been concerned with the movement afforded by the hyoid explicitly (but see Westbury, 1988; Hiiemae et al., 2002).

inferring species' speech capacities, holistic interpretation of those elements – including flexure of the skull base, shape of the hyoid bone, and phylogenetic restructuring of facial morphology – suggest a gradual transition from the apelike vocal anatomy of extinct early human ancestors toward that of modern humans, with *H. erectus* appearing as a likely in-between point. Additionally, no evidence presented on Neanderthal potential speech capacities have convincingly argued that the species' "vowel space was as large of that of modern humans" (Boë et al., 2002a,b). The current state of research tentatively favors Crelin's (1989, p. 19) interpretation, that the "evolution of the [human vocal] tract occurred … quite recently in time."

4. The tongue in speech and speechlike behavior

4.1. A comparative perspective

Possible tongue involvement in great ape articulation is difficult to study via observation alone (Grawunder et al., 2022; Ekström et al., in press), and any procedure typical of phonetics, such as palatography (measurements of tongue position in speech articulation) is not feasibly applicable to non-human subjects. This is crucial, because the likely limitations imposed on species' articulatory capacities from tongue morphology strongly suggests the relevant vocal tract area functions are unattainable by those species by the same means as by humans (Takemoto, 2008; Lieberman, 2012). Indeed, available evidence indicates that a human tongue is necessary for achieving the superior lingual curvature via decompression of the tongue body against the hard palate that characterizes various human speech sounds (e.g., Engwall, 2003) (Table 1). The work by Grawunder et al. (2022) also do not provide strong evidence for any involvement of the tongue in vocalizations, aligning with previous work on nonhuman animal vocalizations (Fitch, 1997, 2000a,b; see also Ekström, 2022a). The resulting "vowel-like space" is rather suggestive of the possibility that chimpanzees shift resonance frequencies down (into an /u:/-like

dispersion) by using the lips, effectively elongating the vocal tract, and narrowing its lip passage (Fant, 1960). For clues to *in-situ* function, however, we may turn to case studies.

In one such study of speechlike utterances by Viki the chimpanzee – raised in a human home and explicitly tutored in speech (Hayes and Hayes, 1951) – one of the authors found that, while Viki had seemingly learned an articulatory gesture roughly corresponding to the lexical form "cup," this sequence was seemingly (as indicated by comparative analysis with a human speaker) realized as a combination of a voiceless fricative, produced in the dorsal oral cavity, and a voiceless bilabial plosive (Ekström, 2023b). That is, humanlike production of the intended sequence or word was seemingly unavailable. Velar plosive [k] (c̲up) requires substantial maneuverability of the tongue body amounting to a brief but complete occlusion of pulmonary airflow in the oral cavity. Thus, both anatomical and acoustic-phonetic evidence suggests that the chimpanzee is precluded from the articulatory finesse exhibited by modern humans. The chimpanzee tongue is likely incapable of such extreme compression as to make possible the rapid tongue-body plosive speech gestures involved in, e.g., [k] and the rapid changes involved in everyday speech. Further, because the chimpanzee oral cavity is larger than that of humans (Lieberman, 2011), the distance to would-be articulatory targets (Perkell, 1996) is also greater.

4.2. Lingual coarticulation: an emergent property

Speech – temporally and in terms of serial organization – is moves between learned gestures. That is, speech gestures are executed continuously, and in natural speech, articulatory organs do not produce any one gesture in isolation. In the words of Farnetani and Recasens (2010, p. 217) "A fundamental and extraordinary characteristic of spoken language … is that the movements of different articulators for the production of successive phonetic segments overlap in time and interact with one another: as a consequence, the

TABLE 1 Portions of the human tongue, and their involvement in a variety of speech gestures.

Sound	Articulation	Phonetic transcription	Example of in-word usage
Close back rounded vowel	Pulmonic airflow through vocal tract, where tongue is positioned close to the hard palate, in the posterior oral cavity ("back").	[u]	"boot"
Close front unrounded vowel	Pulmonic airflow through vocal tract, where tongue is positioned close to the hard palate, in the anterior oral cavity ("front").	[i]	"see"
Voiceless dental alveolar plosive	Temporary occlusion of pulmonic airflow via the tongue tip making contact with front teeth, and tongue sides making contact with anterior and lateral alveolar ridge.	[t]	"tip"
Voiceless velar plosive	Temporary occlusion of pulmonic airflow via the tongue body making contact with the soft palate, molars, and gum ridge.	[k]	"can"
Voiceless dental fricative	Tongue tip or blade making contact with upper teeth, constricting airflow and causing turbulence.	[θ]	"the"

All gestures involve active manipulation of the tongue.

vocal tract configuration at any point in time is influenced by more than one segment." For non-reduced vowels, maneuvers of the tongue root and tongue body necessary for a shift between one vowel to another are likely capped at ~100 ms in humans (Perkell, 1969) – but a vowel reduced from imposition of coarticulatory constrictions (executable without the prior specification of articulatory target positions) may be actualized at still faster rates.

While a number of studies consider the apparent "syntax" of primate calls (Zuberbühler, 2018), such work rarely studies articulatory gestures involved, but the temporal adjacency or connectedness of calls, call types, or acoustic aspects of calls (but see Fitch, 2000a,b; Grawunder et al., 2022). While a central tenant of human speech, and an obligatory component of speech motor behavior, however, the centrality of coarticulation to speech-centric activity is not readily recognized in the broader literature on language evolution. For example, Fitch (2010), in *The Evolution of Language* devotes less than two pages to the subject, arguing that "Coarticulation seems as likely to be an unfortunate byproduct of producing sounds with a massive tongue as a specifically evolved 'feature' of the human vocal tract." In so writing, Fitch addresses claims by Liberman et al. (1967) who argued that coarticulation evolved to meet the demand of perceptual systems. The argument by Liberman and colleagues verges on teleology (explaining phenomena by their function, rather than their ultimate cause), and Fitch (2010) is correct that coarticulation *per se* need not be considered an evolved feature. However, what Fitch calls "an unfortunate byproduct," is likely the exact opposite.

Temporal reduction of speech sounds also results in the compact transmission of information. This is of enormous benefit for human systems of auditory perception and short-term memory, which are not capable of storing infinite amounts of incoming information (Miller, 1956; Shiffrin and Nosofsky, 1994; Cowan, 2001). Indeed, this was the line of argument that initially prompted Liberman et al. (1967) to develop their argument and corresponding "motor theory of speech perception". The supposedly "unfortunate" nature of coarticulation, then, in reality reflects a fortuitous advantage of speech production in general, and of the human tongue, which has evolved the capacity for more articulate speech, compared to that of any other extant animal, in particular. Coarticulation – by selective evolution or by happenstance – allows "chunky" speech perception, the perceptual reduction of acoustic features into consonantal "frames" and vowel "content" (MacNeilage, 1998), effectively concentrating a complex stream of sounds into readily perceivable pseudo units (see also Studdert-Kennedy et al., 1998). Seen from such a perspective, coarticulatory phenomena are suggestive of holistic anatomical-neural coevolution of human speech, rather than a phylogenetic emphasis of one over the other (see Lieberman, 2012, 2017; *cf.* Fitch et al., 2016).

4.3. Learned speech as reaching and grabbing

Tongue position in phonemic articulation abides by principles of motor equivalence, such that articulatory goals are achievable via compensatory motions (Gay et al., 1981). That is, just as a reaching action of the hand and arm is continually adjusted in execution to compensate for perturbations, so may lingual "reaching" of the tongue (toward an articulatory target) be executed similarly (Moayedi et al., 2021). The neurological mapping and maintenance of articulatory targets are likely facilitated via a basal ganglion-motor cortical network (Graybiel, 2005; Enard, 2011; Alm, 2021; Ekström, 2022b), where the cerebellum is responsible for continual adjustment of fine-motor behavior (Paulin, 1993), including those involved in speech (Ackermann, 2008; Alm, 2021). In modern humans, there has been significant phylogenetic development of subcortical structures including the cerebellum (Baizer, 2014; Guevara et al., 2021). This is consistent with the emerging picture in neurolinguistics, that a distributed network, rather than any one or few language center(s), is responsible for linguistic abilities (Lieberman, 2000; Murdoch, 2001; Dronkers et al., 2007; Friederici and Gierhan, 2013; Ekström, 2022b).

In comparison, chimpanzees are evidently capable of mapping hand gestures. The success of Washoe the chimpanzee, who is reported to have learned to use hundreds of signs (Gardner et al., 1989; Jensvold and Gardner, 2000), compared to Viki, who was claimed to have spoken *four* words (Hayes and Hayes, 1951; Ekström, 2023b), suggests that one differentiating factor is the relative ease of manual-gestural mapping, as opposed to lingual-gestural mapping, where flexibility is apparently more extensive for the first. While a variety of gestural origins accounts of language evolution have posited an evolutionary trajectory from "hand to mouth" (Corballis, 2003), such claims are not made here. The emergent picture, rather, is suggestive of the capacity for motor mapping of novel gestures or (possibly) gestural sequences being present already in the common ancestor of humans and chimpanzees (and possible more ancient still; Cartmill and Byrne, 2010), though this capacity was largely limited to manual gestures, possibly stemming from the flexible and intentional use of gestural communication observed in wild chimpanzees (Hobaiter and Byrne, 2014), and comparatively inflexible vocal anatomy (Negus, 1949; Lieberman, 1984, 2012; de Boer and Fitch, 2010; but see Lameira, 2017; Ekström, 2023b). Crucial for human speech, however, was an extension of these capacities, to increasingly flexible lingual anatomy. Future research may seek to understand the nature of neurological systems that underlie the acquisition of novel gestures – manual and lingual – in the human and nonhuman primate. Such a theory would help explicate common psychological origins of spoken and signed languages, as contingent on unified neurological frameworks.

5. Concluding comments

Available evidence from anthropology and acoustic phonetics suggests (1) gradual evolution of central vocal anatomical features crucial to modern human spoken language, possibly beginning with *H. erectus*, and that (2) until such a time that modern human vocal anatomy was achieved, articulate speech was beyond the articulatory capacities now-extinct ancestral hominids. Further, because tongue movements between articulatory targets are learned, a neural mechanism in humans facilitates the mapping of such targets. Comparative research suggests that similar mechanisms may already have been in place in early chimpanzee-like ancestors – but, given species-typical lingual limitations, may have been largely limited to

manual gestures. The evolution of the modern human tongue was an essential element of the evolution of human spoken language.

Author contributions

AE: conceptualization, writing–initial draft, and writing–editing and review. JE: writing–editing and review. All authors contributed to the article and approved the submitted version.

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ackermann, H. (2008). Cerebellar contributions to speech production and speech perception: psycholinguistic and neurobiological perspectives. *Trends Neurosci.* 31, 265–272. doi: 10.1016/j.tins.2008.02.011

Alemseged, Z., Spoor, F., Kimbel, W. H., Bobe, R., Geraads, D., Reed, D., et al. (2006). A juvenile early hominin skeleton from Dikika Ethiopia. *Nature* 443, 296–301. doi: 10.1038/nature05047

Alm, P. A. (2021). The dopamine system and automatization of movement sequences: a review with relevance for speech and stuttering. *Front. Hum. Neurosci.* 15:1880. doi: 10.3389/fnhum.2021.661880

Ankel-Simons, F. (2007). *Primate anatomy: An introduction*. Amsterdam: Elsevier.

Arensburg, B., Tillier, A. M., Vandermeersch, B., Duday, H., Schepartz, L. A., and Rak, Y. (1989). A middle Palaeolithic human hyoid bone. *Nature* 338, 758–760. doi: 10.1038/338758a0

Baizer, J. S. (2014). Unique features of the human brainstem and cerebellum. *Front. Hum. Neurosci.* 8:202. doi: 10.3389/fnhum.2014.00202

Berwick, R. C., Hauser, M. D., and Tattersall, I. (2013). Neanderthal language? Just-so stories take center stage. *Front. Psychol.* 4:671. doi: 10.3389/fpsyg.2013.00671

Boë, L. J., Heim, J. L., and Abry, C. (2002b) The size of the pharynx: an irrelevant parameter for speech emergence and acquisition. Fourth international conference on the evolution of language.

Boë, L. J., Heim, J. L., Honda, K., and Maeda, S. (2002a). The potential Neandertal vowel space was as large as that of modern humans. *J. Phon.* 30, 465–484. doi: 10.1006/jpho.2002.0170

Boë, L. J., Heim, J. L., Honda, K., Maeda, S., Badin, P., and Abry, C. (2007). The vocal tract of newborn humans and Neanderthals: acoustic capabilities and consequences for the debate on the origin of language. A reply to Lieberman (2007a). *J. Phon.* 35, 564–581. doi: 10.1016/j.wocn.2007.06.006

Boë, L. J., Maeda, S., and Heim, J. L. (1999). Neandertal man was not morphologically handicapped for speech. *Evol. Commun.* 3, 49–77. doi: 10.1075/eoc.3.1.05boe

Bosma, J. F. (1975). Anatomic and physiologic development of the speech apparatus. *Nerv. Syst.* 3, 469–481.

Burrows, A. M., Parr, L. A., Durham, E. L., Matthews, L. C., and Smith, T. D. (2014). Human faces are slower than chimpanzee faces. *PLoS One* 9:e110523. doi: 10.1371/journal.pone.0110523

Carlisle, R. C., and Siegel, M. I. (1974). Some problems in the interpretation of Neanderthal speech capabilities: a reply to Lieberman. *Am. Anthropol.* 76, 319–322. doi: 10.1525/aa.1974.76.2.02a00050

Carre, R., Lindblom, B., and MacNeilage, P. (1995). Acoustic factors in the evolution of the human vocal tract. *C. R. Acad Sci II* 320, 471–476.

Cartmill, E. A., and Byrne, R. W. (2010). Semantics of primate gestures: intentional meanings of orangutan gestures. *Anim. Cogn.* 13, 793–804. doi: 10.1007/s10071-010-0328-7

Corballis, M. C. (2003). From hand to mouth: the gestural origins of language. *Stud. Evol. Lang.* 3, 201–218. doi: 10.1093/acprof:oso/9780199244843.003.0011

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav. Brain Sci.* 24, 87–114. doi: 10.1017/S0140525X01003922

Crelin, E. S. (1987). *The human vocal tract: Anatomy, function, development, and evolution*. New York: Vantage Press.

Crelin, E. S. (1989). The skulls of our ancestors: Implications regarding speech, language, and conceptual thought evolution. *J. Voice* 3, 18–23.

D'Anastasio, R., Wroe, S., Tuniz, C., Mancini, L., Cesana, D. T., Dreossi, D., et al. (2013). Micro-biomechanics of the Kebara 2 hyoid and its implications for speech in Neanderthals. *PLoS One* 8:e82261. doi: 10.1371/journal.pone.0082261

Dart, R. A. (1925). Australopithecus africanus: the man-ape of South Africa. *Nature* 115, 195–199. doi: 10.1038/115195a0

de Boer, B., and Fitch, T. W. (2010). Computer models of vocal tract evolution: an overview and critique. *Adapt. Behav.* 18, 36–47. doi: 10.1177/1059712309350972

Dediu, D., and Levinson, S. C. (2013). On the antiquity of language: the reinterpretation of Neandertal linguistic capacities and its consequences. *Front. Psychol.* 4:397. doi: 10.3389/fpsyg.2013.00397

DeGusta, D., Gilbert, W. H., and Turner, S. P. (1999). Hypoglossal canal size and hominid speech. *Proc. Natl. Acad. Sci.* 96, 1800–1804. doi: 10.1073/pnas.96.4.1800

Dronkers, N. F., Plaisant, O., Iba-Zizen, M. T., and Cabanis, E. A. (2007). Paul Broca's historic cases: high resolution MR imaging of the brains of Leborgne and Lelong. *Brain* 130, 1432–1441. doi: 10.1093/brain/awm042

Ekström, A. G. (2022a). Ape vowel-like sounds remain elusive: a comment on Grawunder et al. (2022). *Int. J. Primatol.* 44, 237–239. doi: 10.1007/s10764-022-00335-6

Ekström, A. G. (2022b). Motor constellation theory: a model of infants' phonological development. *Front. Psychol.* 13:6894. doi: 10.3389/fpsyg.2022.996894

Ekström, A. G. (2023a). *Not 'speech-ready' after all: a critical review of nonhuman primate phonetic potential*. [Preprint]. doi: 10.31219/osf.io/hd48k

Ekström, A. G. (2023b). Viki's first words: a comparative phonetics case study. *Int. J. Primatol.* 44, 249–253. doi: 10.1007/s10764-023-00350-1

Ekström, A. G., Moran, S., Sundberg, J., and Lameira, A. (in press). PREQUEL: Supervised phonetic approaches to analyses of great ape quasi-vowels. *Proceedings of 20th International Congress of Phonetic Sciences (ICPhS 2023)*.

Enard, W. (2011). FOXP2 and the role of cortico-basal ganglia circuits in speech and language evolution. *Curr. Opin. Neurobiol.* 21, 415–424. doi: 10.1016/j.conb.2011.04.008

Engwall, O. (2003). Combining MRI, EMA and EPG measurements in a three-dimensional tongue model. *Speech Comm.* 41, 303–329. doi: 10.1016/S0167-6393(02)00132-2

Falk, D. (1975). Comparative anatomy of the larynx in man and the chimpanzee: implications for language in Neanderthal. *Am. J. Phys. Anthropol.* 43, 123–132. doi: 10.1002/ajpa.1330430116

Fant, G. (1960). *The acoustic theory of speech production*. Mouton: The Hague.

Farnetani, E., and Recasens, D. (2010). "Coarticulation and connected speech processes" in . eds. W. J. Hardcastle, J. Laver and F. E. Gibbon (Hoboken, NJ, Wiley-Blackwell: The handbook of phonetic sciences), 316–352.

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. doi: 10.1121/1.421048

Fitch, W. T. (2000a). The phonetic potential of nonhuman vocal tracts: comparative cineradiographic observations of vocalizing animals. *Phonetica* 57, 205–218. doi: 10.1159/000028474

Fitch, W. T. (2000b). The evolution of speech: a comparative review. *Trends Cogn. Sci.* 4, 258–267. doi: 10.1016/S1364-6613(00)01494-7

Fitch, W. T. (2009). "Fossil cues to the evolution of speech," in *The cradle of language*. eds. R. Botha and C. Knight (Oxford University Press), 112–134.

Fitch, W. T. (2010). *The evolution of language*. Cambridge: Cambridge University Press.

Fitch, W. T., De Boer, B., Mathur, N., and Ghazanfar, A. A. (2016). Monkey vocal tracts are speech-ready. *Sci. Adv.* 2:e1600723. doi: 10.1126/sciadv.1600723

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. doi: 10.1121/1.427148

Fitch, T. W., and Reby, D. (2001). The descended larynx is not uniquely human. *Proc. Royal Soc. London B Biol. Sci.* 268, 1669–1675. doi: 10.1098/rspb.2001.1704

Fleagle, J. G. (2013). *Primate adaptation and evolution*. Cambridge: Academic press.

Frayer, D. W. (2017). "Talking hyoids and talking Neanderthals," in *Human Paleontology and Prehistory: Contributions in Honor of Yoel Rak*. eds. A. Marom and E. Hovers (Springer), 233–237.

Friederici, A. D., and Gierhan, S. M. (2013). The language network. *Curr. Opin. Neurobiol.* 23, 250–254. doi: 10.1016/j.conb.2012.10.002

Gardner, R. A., Gardner, B. T., and Van Cantfort, T. E. (1989). *Teaching sign language to chimpanzees*. Albany, NY: Suny Press.

Gay, T., Lindblom, B., and Lubker, J. (1981). Production of bite-block vowels: acoustic equivalence by selective compensation. *J. Acoust. Soc. Am.* 69, 802–810. doi: 10.1121/1.385591

Gilbert, R. J., Napadow, V. J., Gaige, T. A., and Wedeen, V. J. (2007). Anatomical basis of lingual hydrostatic deformation. *J. Exp. Biol.* 210, 4069–4082. doi: 10.1242/jeb.007096

Grawunder, S., Uomini, N., Samuni, L., Bortolato, T., Girard-Buttoz, C., Wittig, R. M., et al. (2022). Chimpanzee vowel-like sounds and voice quality suggest formant space expansion through the hominoid lineage. *Philos. Trans. R. Soc. B* 377:20200455. doi: 10.1098/rstb.2020.0455

Graybiel, A. M. (2005). The basal ganglia: learning new tricks and loving it. *Curr. Opin. Neurobiol.* 15, 638–644. doi: 10.1016/j.conb.2005.10.006

Guevara, E. E., Hopkins, W. D., Hof, P. R., Ely, J. J., Bradley, B. J., and Sherwood, C. C. (2021). Comparative analysis reveals distinctive epigenetic features of the human cerebellum. *PLoS Genet.* 17:e1009506. doi: 10.1371/journal.pgen.1009506

Hayes, K. J., and Hayes, C. (1951). The intellectual development of a home-raised chimpanzee. *Proc. Am. Philos. Soc.* 95, 105–109.

Heim, J. L. (1989). La nouvelle reconstitution du crâne Néandertalien de La Chapelle-aux-Saints: méthode et résultats. *Bulletins et Memoires de la Societe d'Anthropologie de Paris* 1, 95–117. doi: 10.3406/bmsap.1989.1702

Hiiemae, K. M., Palmer, J. B., Medicis, S. W., Hegener, J., Jackson, B. S., and Lieberman, D. E. (2002). Hyoid and tongue surface movements in speaking and eating. *Arch. Oral Biol.* 47, 11–27. doi: 10.1016/S0003-9969(01)00092-9

Hobaiter, C., and Byrne, R. W. (2014). The meanings of chimpanzee gestures. *Curr. Biol.* 24, 1596–1600. doi: 10.1016/j.cub.2014.05.066

Ishwar, . (2020). "Cardinal vowel tongue position.png" [Image]. Wikipedia Commons. Available at: https://commons.wikimedia.org/wiki/File:Cardinal_vowel_tongue_position.png (Accessed January 22, 2023).

Iwasaki, S. I., Yoshimura, K., Shindo, J., and Kageyama, I. (2019). Comparative morphology of the primate tongue. *Ann. Anat. Anatomischer Anzeiger* 223, 19–31. doi: 10.1016/j.aanat.2019.01.008

Iwasaki, M. L. A., and Gardner, R. A. (2000). Interactive use of sign language by cross-fostered chimpanzees (*Pan troglodytes*). *J. Comp. Psychol.* 114, 335–346. doi: 10.1037/0735-7036.114.4.335

Jungers, W. L., Pokempner, A. A., Kay, R. F., and Cartmill, M. (2003). Hypoglossal canal size in living hominoids and the evolution of human speech. *Hum. Biol.* 75, 473–484. doi: 10.1353/hub.2003.0057

Kay, R. F., Cartmill, M., and Balow, M. (1998). The hypoglossal canal and the origin of human vocal behavior. *Proc. Natl. Acad. Sci.* 95, 5417–5419. doi: 10.1073/pnas.95.9.5417

Laitman, J. T., and Heimbuch, R. C. (1982). The basicranium of Plio-Pleistocene hominids as an indicator of their upper respiratory systems. *Am. J. Phys. Anthropol.* 59, 323–343. doi: 10.1002/ajpa.1330590315

Laitman, J. T., Heimbuch, R. C., and Crelin, E. S. (1978). Developmental change in a basicranial line and its relationship to the upper respiratory system in living primates. *Am. J. Anat.* 152, 467–482. doi: 10.1002/aja.1001520403

Lameira, A. R. (2017). Bidding evidence for primate vocal learning and the cultural substrates for speech evolution. *Neurosci. Biobehav. Rev.* 83, 429–439. doi: 10.1016/j.neubiorev.2017.09.021

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol. Rev.* 74, 431–461. doi: 10.1037/h0020279

Lieberman, P. (1984). *The biology and evolution of language*. Cambridge, MA: Harvard University Press.

Lieberman, P. (1999). Silver-tongued Neandertals? *Science* 283:175. doi: 10.1126/science.283.5399.175b

Lieberman, P. (2000). *Human language and our reptilian brain: The subcortical bases of speech, syntax, and thought*. Cambridge, MA: Harvard University Press.

Lieberman, P. (2007a). The evolution of human speech: its anatomical and neural bases. *Curr. Anthropol.* 48, 39–66. doi: 10.1086/509092

Lieberman, P. (2007b). Current views on Neanderthal speech capabilities: a reply to Boe et al.(2002). *J. Phon.* 35, 552–563. doi: 10.1016/j.wocn.2005.07.002

Lieberman, D. (2011). *The evolution of the human head*. Cambridge, MA: Harvard University Press.

Lieberman, P. (2012). Vocal tract anatomy and the neural bases of talking. *J. Phon.* 40, 608–622. doi: 10.1016/j.wocn.2012.04.001

Lieberman, P. (2017). Comment on "monkey vocal tracts are speech-ready". *Sci. Adv.* 3:e1700442. doi: 10.1126/sciadv.1700442

Lieberman, P., and Crelin, E. S. (1971). On the speech of Neanderthal man. *Janua Linguarum* 76

Lieberman, P., and Crelin, E. S. (1974). Speech and Neanderthal man: a reply to Carlisle and Siegel. *Am. Anthropol.* 76, 323–325. doi: 10.1525/aa.1974.76.2.02a00060

Lieberman, P., Crelin, E. S., and Klatt, D. H. (1972). Phonetic ability and related anatomy of the newborn and adult human, Neanderthal man, and the chimpanzee. *Am. Anthropol.* 74, 287–307. doi: 10.1525/aa.1972.74.3.02a00020

Lieberman, P. H., Klatt, D. H., and Wilson, W. H. (1969). Vocal tract limitations on the vowel repertoires of rhesus monkey and other nonhuman primates. *Science* 164, 1185–1187. doi: 10.1126/science.164.3884.1185

Lieberman, P., Laitman, J. T., Reidenberg, J. S., and Gannon, P. J. (1992). The anatomy, physiology, acoustics and perception of speech: essential elements in analysis of the evolution of human speech. *J. Hum. Evol.* 23, 447–467. doi: 10.1016/0047-2484(92)90046-C

Lieberman, P., Laitman, J. T., Reidenberg, J. S., Landahl, K., and Gannon, P. J. (1989). Folk physiology and talking hyoids. *Nature* 342, 486–487. doi: 10.1038/342486a0

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. doi: 10.1006/jhev.1998.0287

Lieberman, D. E., McCarthy, R. C., Hiiemae, K. M., and Palmer, J. B. (2001). Ontogeny of postnatal hyoid and larynx descent in humans. *Arch. Oral Biol.* 46, 117–128.

Machida, H., Perkins, E., and Giacometti, L. (1967). The anatomical and histochemical properties of the tongue of primates. *Folia Primatol.* 5, 264–279. doi: 10.1159/000161951

MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behav. Brain Sci.* 21, 499–511. doi: 10.1017/S0140525X98001265

Maeda, S. (1990). "Compensatory articulation during speech: evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model," *Speech production and speech modelling*. eds. W. H. Hardcastle and A. Marchal (Springer), 131–149.

Martínez, I., Arsuaga, J. L., Quam, R., Carretero, J. M., Gracia, A., and Rodríguez, L. (2008). Human hyoid bones from the middle Pleistocene site of the Sima de los Huesos (sierra de Atapuerca, Spain). *J. Hum. Evol.* 54, 118–124. doi: 10.1016/j.jhevol.2007.07.006

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* 63, 81–97. doi: 10.1037/h0043158

Moayedi, Y., Michlig, S., Park, M., Koch, A., and Lumpkin, E. A. (2021). Somatosensory innervation of healthy human oral tissues. *J. Comp. Neurol.* 529, 3046–3061. doi: 10.1002/cne.25148

Murdoch, B. E. (2001). Subcortical brain mechanisms in speech and language. *Folia Phoniatr. Logop.* 53, 233–251. doi: 10.1159/000052679

Negus, V. E. (1949). *Comparative anatomy and physiology of the larynx*. Heinemann.

Nishimura, T. (2005). Developmental changes in the shape of the supralaryngeal vocal tract in chimpanzees. *Am. J. Phys. Anthropol.* 126, 193–204. doi: 10.1002/ajpa.20112

Osvath, M., and Gärdenfors, P. (2005). Oldowan culture and the evolution of anticipatory cognition. *Lund Univ. Cogn. Stud.* 122, 1–16.

Paulin, M. G. (1993). The role of the cerebellum in motor control and perception. *Brain Behav. Evol.* 41, 39–50. doi: 10.1159/000113822

Perkell, J. S. (1969). Physiology of speech production: results and implication of quantitative cineradiographic study. *Monograph* 53

Perkell, J. S. (1996). Properties of the tongue help to define vowel categories: hypotheses based on physiologically-oriented modeling. *J. Phon.* 24, 3–22. doi: 10.1006/jpho.1996.0002

Sanders, I., Mu, L., Amirali, A., Su, H., and Sobotka, S. (2013). The human tongue slows down to speak: muscle fibers of the human tongue. *Anat. Rec.* 296, 1615–1627. doi: 10.1002/ar.22755

Shiffrin, R. M., and Nosofsky, R. M. (1994). Seven plus or minus two: a commentary on capacity limitation. *Psychol. Rev.* 101, 357–361. doi: 10.1037/0033-295X.101.2.357

Smith, K. K., and Kier, W. M. (1989). Trunks, tongues, and tentacles: moving with skeletons of muscle. *Am. Sci.* 77, 28–35.

Steele, J., Clegg, M., and Martelli, S. (2013). Comparative morphology of the hominin and african ape hyoid bone, a possible marker of the evolution of speech. *Hum. Biol.* 85, 639–672.

Stevens, K. N. (1989). On the quantal nature of speech. *J. Phon.* 17, 3–45. doi: 10.1016/S0095-4470(19)31520-7

Studdert-Kennedy, M., Hurford, J. R., and Knight, C. (1998). "The particulate origins of language generativity: from syllable to gesture" in *Approaches to the evolution of language: Social and cognitive bases*. ed. J. Hurford (Cambridge: Cambridge University Press), 202–221.

Swindler, D. R., and Wood, C. D. (1982). *An atlas of primate gross anatomy: Baboon, chimpanzee, and man*. Malabar: Krieger Publishing Co.

Takemoto, H. (2008). Morphological analyses and 3D modeling of the tongue musculature of the chimpanzee (*Pan troglodytes*). *Am. J. Primatol.* 70, 966–975. doi: 10.1002/ajp.20589

Weissengruber, G. E., Forstenpointner, G., Peters, G., Kübber-Heiss, A., and Fitch, W. T. (2002). Hyoid apparatus and pharynx in the lion (*Panthera leo*), jaguar (*Panthera onca*), tiger (*Panthera tigris*), cheetah (*Acinonyx jubatus*) and domestic cat (*Felis silvestris* f. catus). *J. Anat.* 201, 195–209. doi: 10.1046/j.1469-7580.2002.00088.x

Westbury, J. R. (1988). Mandible and hyoid bone movements during speech. *J. Speech Lang. Hear. Res.* 31, 405–416. doi: 10.1044/jshr.3103.405

White, T. D., Asfaw, B., Beyene, Y., Haile-Selassie, Y., Lovejoy, C. O., Suwa, G., et al. (2009). Ardipithecus ramidus and the paleobiology of early hominids. *Science* 326, 64–86. doi: 10.1126/science.1175802

Wollstonecroft, M. M. (2011). Investigating the role of food processing in human evolution: a niche construction approach. *Archaeol. Anthropol. Sci.* 3, 141–150. doi: 10.1007/s12520-011-0062-3

Wrangham, R. (2009). *Catching fire: How cooking made us human*. New York: Basic books.

Zuberbühler, K. (2018). Combinatorial capacities in primates. *Curr. Opin. Behav. Sci.* 21, 161–169. doi: 10.1016/j.cobeha.2018.03.015