



# Correlates of Vocal Tract Evolution in Late Pliocene and Pleistocene Hominins

Axel G. Ekström<sup>1,2</sup> · Peter Gärdenfors<sup>3,4</sup> · William D. Snyder<sup>5,6</sup> · Daniel Friedrichs<sup>2,7</sup> · Robert C. McCarthy<sup>8</sup> · Melina Tsapos<sup>3</sup> · Claudio Tennie<sup>6</sup> · David S. Strait<sup>4,9,10</sup> · Jens Edlund<sup>1</sup> · Steven Moran<sup>2,7,11</sup>

Accepted: 20 February 2025 / Published online: 17 April 2025  
© The Author(s) 2025

## Abstract

Despite decades of research on the emergence of human speech capacities, an integrative account consistent with hominin evolution remains lacking. We review paleoanthropological and archaeological findings in search of a timeline for the emergence of modern human articulatory morphological features. Our synthesis shows that several behavioral innovations coincide with morphological changes to the would-be speech articulators. We find that significant reductions of the mandible and masticatory muscles and vocal tract anatomy coincide in the hominin fossil record with the incorporation of processed and (ultimately) cooked food, the appearance and development of rudimentary stone tools, increases in brain size, and likely changes to social life and organization. Many changes are likely mutually reinforcing; for example, gracilization of the hominin mandible may have been maintainable in the lineage because food processing had already been outsourced to the hands and stone tools, reducing selection pressures for robust mandibles in the process. We highlight correlates of the evolution of craniofacial and vocal tract features in the hominin lineage and outline a timeline by which our ancestors became ‘pre-adapted’ for the evolution of fully modern human speech.

**Keywords** Evolution of speech · Biological anthropology · Articulatory phonetics · Cognitive evolution · Paleoanthropology · Cooking hypothesis

## Toward an Integrative Account of Hominin Vocal Tract Evolution

Much remains unknown about the selection pressures and sequence of events that facilitated the evolution of speech in hominins. Some aspects of these events may be traced via studies of comparative facial morphology. Humans have shallow, vertical faces, mandibles and teeth reduced in size compared to apes and other hominin species (Wrangham, 2009; D. Lieberman, 2011; Puts et al., 2012; Lordkipanidze et al.,

---

Extended author information available on the last page of the article

2013; Katz et al., 2017; von Cramon-Taubadel, 2017; Lacruz et al., 2019; Zollikofer et al., 2022), and a tongue and supralaryngeal vocal tract (SVT) remarkably distinct from those of extant non-human primates (Negus, 1949; P. Lieberman, 1984, 2012; Crelin, 1987; Studdert-Kennedy, 1998; Takemoto, 2008; de Boer & Fitch, 2010; D. Lieberman, 2011; Ekström & Edlund, 2023a). The result of these restructurings of *Homo sapiens*’ craniofacial anatomy represents the creation of one of the most derived and most phonetically efficient in existence (Carré et al., 2017; Lindblom, 1983; MacNeilage, 1998).

Traditionally, literature on the subject holds that non-human primate phonetic capacities allow for a rudimentary system of speech. In this view, the fact that no such system is borne out in nature possibly reflect neural (P. Lieberman et al., 1972; Lieberman et al., 2002, 2012, 2017; MacNeilage, 1998; Jürgens, 2002; Lameira, 2017; Fitch et al., 2016; Belyk and Brown, 2017; Brown et al., 2021) and/or genetic limitations (Enard et al., 2002; Trinkaus, 2007; Fisher and Scharff, 2009). The absence of rudimentary speech in non-human primates is thus taken as evidence that other pressures drove the early evolution of speech articulators, while less articulate “early speech” may have enacted an independent selection pressure on the evolution of fully articulate speech anatomy in later human evolution, counteracting the negative side effects of increased choking risk (Negus, 1949; P. Lieberman, 1984, 2012, 2017).

However, recent advances in primate vocal control and production have disputed a number of preconceived notions. Chimpanzees exhibit several prerequisites of spoken language such as lateralization of the temporal lobes (Gannon et al., 1998) and frontal lobes (Cantalupo & Hopkins, 2001; see also Amiez et al., 2023), and can in theory produce a significant range of consonant distinctions including labial consonants such as [m], [b], and [p] (P. Ekström, 2023; Ekström et al., 2024a; Lameira & Moran, 2023; Lameira et al., 2013; Lieberman et al., 1972), dental consonants (see P. Lieberman et al., 1992), and a limited range of vowel sounds (P. Fitch et al., 2016; Lieberman et al., 1969, 1972). Primate vocal behavior exhibits a number of important features consistent with speech behavior (Table 1). As such, there is a need for an integrative account of the emergence of speech capacities that is consistent with current paleoanthropological and archaeological science. Here, we highlight a variety of changes to the hominin vocal tract, and place the emergence of these features on a common timeline with other findings.

## Basics of Speech Production

Human articulate speech is a combined respiratory, phonatory, and supralaryngeal articulatory series of actions. Pulmonic airflow from the respiratory organs causes controlled vibration of the vocal folds in the larynx; the resulting voice “source” is “filtered” (Fant, 1960) by rapid and voluntary alternations between constrictions on airflow inside the SVT, resulting in controlled manipulation of resulting frequencies. For example, in producing the close front unrounded vowel [i] (the vowel in “see”), the tongue is positioned close to the hard palate in the anterior oral cavity. This creates a narrow constriction in the oral tract, shifting up the second formant

**Table 1** Speech-like behavior observed in great apes. Studies on non-hominids are not included. Here, we are interested in whether these behaviors are possible at all; how such behaviors were learned (e.g., through enculturation) is not emphasized (cf. e.g., Lameira, 2017; Motes-Rodrigo & Tennie, 2021)

Prerequisite	Behavior	Species	Reference
<i>Breath and voice control</i>	Volitional breath control during play with wind instruments	<i>Gorilla gorilla</i>	Perlman et al. (2012)
	Voluntary control to produce labial and lingual-labial fricative sounds	<i>Gorilla gorilla</i>	Perlman and Clark (2015)
	Inhalation to retrieve food items, exhalation to elevate a ball inside clear cylinder	<i>Pan troglodytes</i> <i>Pan paniscus</i>	Schwob (2017)
	Volitional voicing and glottal fricative sounds	<i>Gorilla gorilla</i>	Perlman and Clark (2015)
	Active voicing through a membranophone	<i>Pongo</i> <sup>a</sup> <i>Pongo abelii</i>	Lameira and Shumaker (2019)
“Speech-like” rhythms	Volitional utterances of the phonetic form [mama] (“mama”)	<i>Pan troglodytes</i>	Ekström et al. (2024a)
	Open-close mandibular cycles within typical frequency window of conversational speech	<i>Pongo pygmaeus</i>	Lameira et al. (2015)
	Open-close mandibular cycles within typical frequency window of conversational speech	<i>Pan troglodytes</i>	Pereira et al. (2020)
	A “vowel-like space” delimited by [a]-like and [u]-like extremes	<i>Pan troglodytes</i>	Grawunder et al. (2022)
	[u]-like space observed	<i>Pongo pygmaeus</i>	Ekström et al. (2023)
<i>Call modification</i>	Individual modulates vocal output based on context	<i>Pan paniscus</i>	Tagliatela et al. (2003)
	Call structural convergence between neighboring populations	<i>Pan troglodytes</i>	Crockford et al. (2004)
	Modification of duration and number of whistles	<i>Pongo</i> <sup>a</sup>	Wich et al. (2009)
	Population-specific calls independent of genetic variation among populations	<i>Pongo pygmaeus</i> <i>Pongo abelii</i>	Wich et al. (2012)
	Hand-assisted kiss-squeaks	<i>Pongo pygmaeus</i>	de Boer et al. (2015)
<i>Novel calls</i>	Greater vocal innovation in high-density population individual (new call types typically short-lived)	<i>Pongo pygmaeus</i> , <i>Pongo abelii</i>	Lameira et al. (2022)
	At least two learned utterances, “cup” and “papa”	<i>Pongo</i> <sup>b</sup>	Furness (1916)
	A juvenile learned to reproduce words	<i>Pan troglodytes</i>	Hayes and Hayes (1951)
	Species-atypical attention-getting vocalizations	<i>Pan troglodytes</i>	Hopkins et al. (2007)

**Table 1** (continued)

Prerequisite	Behavior	Species	Reference
<i>Larynx-SVT coupling</i>	Acquired human whistle	<i>Pongo</i> <sup>d</sup>	Wich et al. (2009)
	Mothers produce novel “come hither” calls to infants, the “harmonic uuh”	<i>Pongo abelii</i>	Wich et al. (2012)
	Offspring learn attention getting calls from mothers	<i>Pan troglodytes</i>	Tagliatela et al. (2012)
	Novel attention-getting vocalization through operant conditioning	<i>Pan troglodytes</i>	Russell et al. (2013)
	Ostensibly novel vocalization, the “wookie”	<i>Pongo</i> <sup>d</sup>	Lameira et al. (2016)
	A novel attention-getting vocalization	<i>Gorilla gorilla</i>	Salmi et al. (2022a)
	Utterances of “cup”, “papa” likely correspond to pharyngeal fricative (“c”) and labial plosive (“p”) consonant sounds not observed in vocal repertoires	<i>Pan troglodytes</i>	Ekström (2023)
	Affirms conclusions by Ekström (2023), retracted tongue position inferred	<i>Pan troglodytes</i>	Shofner (2024)
	Two individuals learned to reproduce the phonetic form “mama”	<i>Pan troglodytes</i>	Ekström et al. (2024a)
	Combination of labial fricative and vowel-like utterance with rising F <sub>2</sub> ; “Kiss-squeaks + rolling calls”	<i>Pongo abelii</i>	Lameira and Hardus (2023)
<i>Compositionality</i>	Disyllabic utterances	<i>Pan troglodytes</i>	Ekström et al., (2024a)
	Combination of pant hoots and food calls	<i>Pan troglodytes</i>	Leroux et al. (2021)
	Combination of “Alarm-huu + waa-bark”	<i>Pan troglodytes</i>	Leroux et al. (2023)
<i>Audition</i>	Vocal tract normalization (ignored speaker sex and recognized the same vowel appropriately)	<i>Pan troglodytes</i>	Kojima and Kiritani (1989)
	Discrimination of (some) consonant phonemes, below-human-level performance overall	<i>Pan troglodytes</i>	Kojima et al. (1989)
	Above-human-level sensitivity for frequencies above 8 kHz, less-than-human sensitivity for frequencies at 2-to-4 kHz	<i>Pan troglodytes</i>	Kojima (1990)
	Synthetic speech recognition at above-chance level	<i>Pan troglodytes</i>	Heimbauer et al. (2011)
<i>Recognition</i>	Top-down processes facilitate speech perception when signal is presented in disrupted form	<i>Pan troglodytes</i>	Heimbauer et al. (2021)
	Matching conspecific vocalizations to faces	<i>Pan troglodytes</i>	Izumi and Kojima (2004)
	Better recognition for natural than synthetic speech; both above chance level	<i>Pan paniscus</i>	Lahiff et al. (2022)

**Table 1** (continued)

Prerequisite	Behavior	Species	Reference
<i>Neural anatomy</i>	Recognition of familiar human voices	<i>Gorilla gorilla</i>	Salmi et al. (2022b)
	Left planum temporale larger in 17 of 18 brains	<i>Pan troglodytes</i>	Gannon et al. (1998)
	Frontal lobe asymmetry (of BA44) consistent with left-hemispheric dominance	<i>Pan troglodytes</i>	Cantalupo and Hopkins (2001)
		<i>Pan paniscus</i>	
		<i>Gorilla gorilla</i>	
	Left-hemispheric inferior frontal gyrus activity associated with vocal production	<i>Pan troglodytes</i>	Tagliabata et al. (2008)
	Selective right-lateralized activity in posterior temporal lobe in response to some (not all) calls	<i>Pan troglodytes</i>	Tagliabata et al. (2009)
	Voluntary vocalizations associated with grey matter increases in ventrolateral prefrontal and dorsal premotor cortices	<i>Pan troglodytes</i>	Bianchi et al. (2016)

<sup>a</sup> Hybrid<sup>b</sup> Species undefined; work preceded modern species division

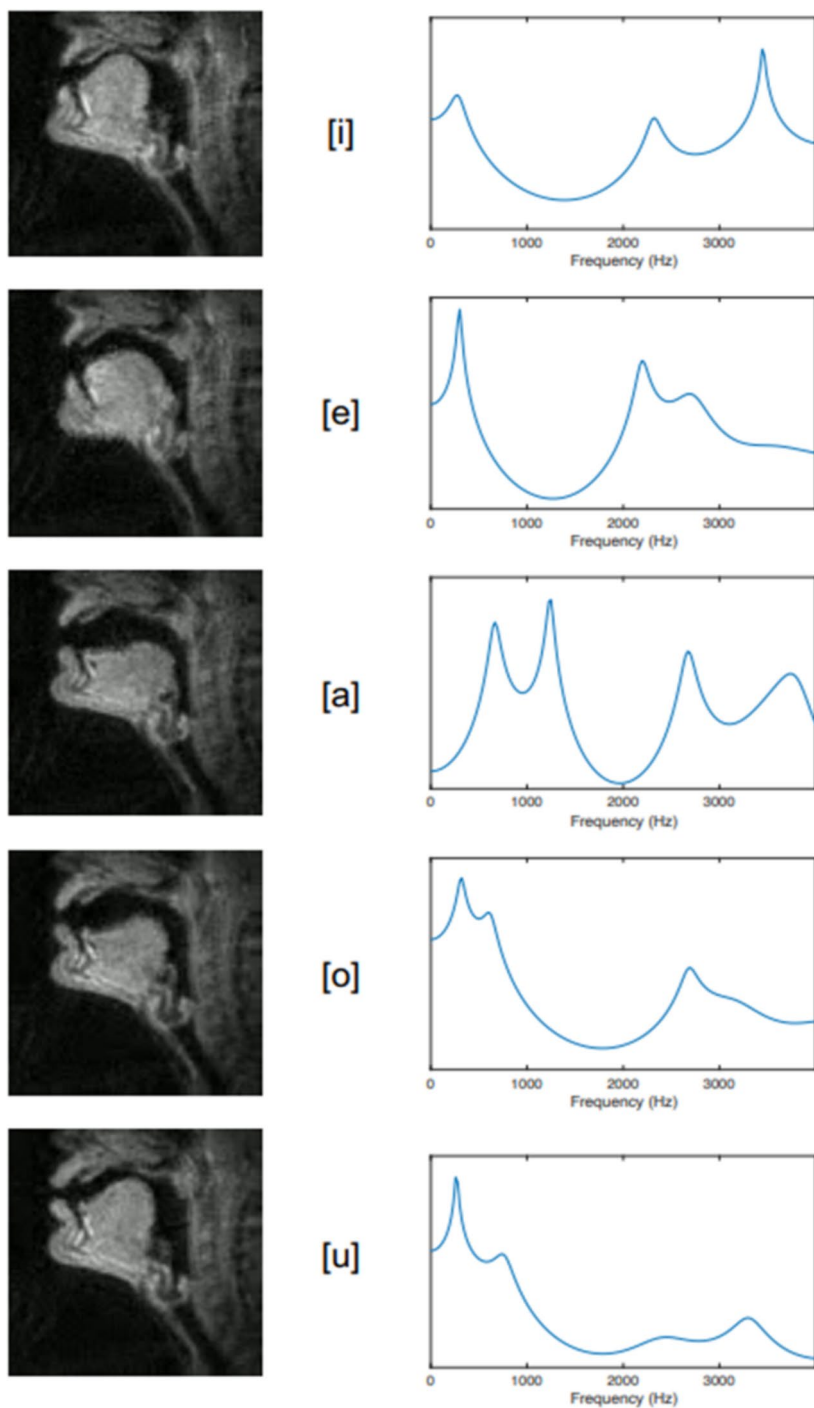
frequency. Observations by Stevens (1989) showed that regions of articulatory space are stable with regard to the acoustic signal produced; for regions corresponding to “stable” vowels [a], [i], and [u], vowel quality can be achieved even when articulation is imprecise (Stevens, 1989; but see also Diehl, 1989, 2008). These vowels are also called point vowels, referring to the extremity of their articulations, as pictured in the International Phonetics Association (IPA) vowel chart (Fig. 1 and Fig. 2).

Other speech sounds are produced in different ways. For example, the voiceless velar plosive [k] (the first consonant in “cat”) involves a brief occlusion of pulmonary airflow using the tongue body followed by a rapid release burst. When coarticulated (produced as part of a sequence of speech sounds, as in everyday speaking), for example as part of a [VkV] (vowel-/k/-vowel) utterance (e.g., “iki”), the consonant imposes a brief but total suspension of voice. In humans, the ritualizing and socially deliberated and negotiated use and reuse of such intra-oral gestures form the bases for phonological systems in all the world’s spoken languages (Fant, 1960; Liljen-crants & Lindblom, 1972; Moran & McCloy, 2019; Stevens, 1989). The consistent and reliable acquisition of such articulatory gestures by human infants in infancy and toddlerhood (Ekström, 2022; Kuhl & Meltzoff, 1996; Lindblom & MacNeilage, 2011; Vihman, 2014) represents a rapid transition from non-speech to speech, an expansion of combinatorial complexity far outweighing any other in nature (Corbal-lis, 2002; Doupe & Kuhl, 1999).

The articulatory, acoustic, and perceptual structure of human speech is “chunky” by fortuitous design, largely organized into series of syllables consisting of vowel “content” couched in consonantal “frames” (MacNeilage, 1998). Articulatory gestures<sup>1</sup> between two consecutive phonemes during speech production is not a one-parameter trajectory. Kinematics of a consonant–vowel syllable are more accurately described via a combinatorial score specifying a series of actions to be carried out by articulatory, phonatory, and respiratory organs. Even seemingly simple utterances are essentially multi-channel events. Take, for example, the syllable [ku] (“koo”). The velar plosive [k] is produced via the brief-but-complete occlusion (called “stop”) of airflow in the oral tract by the tongue body against the hard palate (an “occlusion”), which is then released in a “plosion” of energy. However, in [ku], lip rounding for [u] is observed progressing while closure is being executed, and the observable formant dispersion suggests that the tongue is in a back position at the moment of release. For all articulators to be in position for [u] shortly after the release of [k], rounding of the tongue and other movements must be initiated well in advance. This sequence demonstrates that movements of any two adjacent speech sounds always overlap in time, a universal principle of coarticulation.

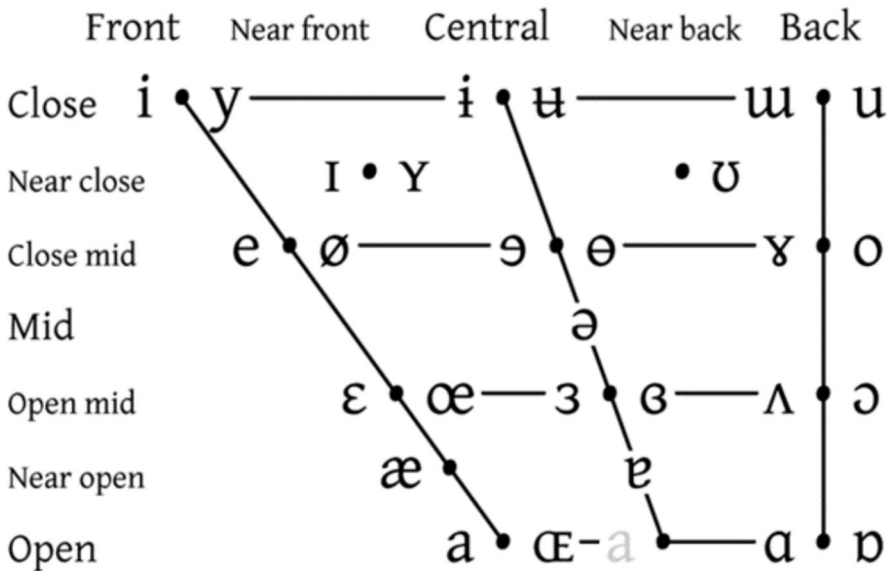
Speech, thus, is not only a matter of control, but also the physiological elements that allows for fine-grain orientation and maneuverability inherent to continuous speech (Ekström & Edlund, 2023a; Lindblom et al., 2009; Liu et al., 2022; Öhman, 1967; Studdert-Kennedy, 1998). Research on listener perception of stop consonants (e.g., [p k t], the first consonants in “pat”, “cat”, and “tat,” respectively) illustrates

<sup>1</sup> Note that throughout this text, we use the terminology of phonetics, such that *gesture* refers to movements of articulators affecting the speech signal, not to manual movements or facial expressions.



**Fig. 1** IPA Vowel chart. The front-to-back and close-to-open dimensions denote stereotyped tongue position and degree of stricture, respectively

## VOWELS



Vowels at right & left of bullets are rounded & unrounded.

**Fig. 2** Articulatory configurations (left), and filter functions (right) for vowel tokens /i e a o u/ as produced by an adult male German speaker

that coordinated speech activity serves as identifying cues, with formant transition patterns indicating subsequent consonants (Delattre et al., 1955; Dorman et al., 1977; Kewley-Port, 1982; Liberman et al., 1967). Formant transitions in [dV] utterances illustrate this point. Like [p k t], [d] is a stop consonant. In speaking [di] (“dee”),  $F_2$  exhibits a telltale upward shift, while for [du] (“doo”) the transition is in the opposite direction (the transition for  $F_1$  is the same for both syllables). However, human vocal anatomy did not spring into existence simply “for purposes of speech” (Negus, 1949): relevant morphology evolved from pre-existing anatomical structures which themselves evolved for other functional and behavioral roles.

Evolution equipped modern humans with the capacity to produce sounds in isolation, in combination, and at variable rates (requiring anatomical evolution), the capacity for planning the production of subsequent sounds (dependent on neural evolutionary changes), and the “cultural consciousness” to enable the persistent, cumulative, and socially negotiated use of syllabic and phonemic vocal communication. To explain the richness, distinctiveness, and full extent of human speech, a combined account of anatomical, neural, and cultural evolution is not only desirable, but necessary. To this end, we trace the correlates of the evolving articulatory complex in human ancestors. Our account connects evolutionary changes in the articulatory complex and vocal tract (P. Lieberman et al., 1969, 1972, 1992; P. Lieberman, 1984, 2012; Carré et al., 1995; Nishimura, 2005; Takemoto, 2008; de Boer & Fitch,



2010; D. Lieberman, 2011; Ekström & Edlund, 2023a) to the advent of mechanical food processing (Semaw et al., 1997; Panger et al., 2002; Gott, 2002; Wrangham & Conklin-Brittain, 2003; Wrangham, 2009; Zink & D. Lieberman, 2016; Snyder et al., 2022), the cognitive faculties that made them possible (Gärdenfors & Högborg, 2017; Gärdenfors & Lombard, 2018; Lombard & Gärdenfors, 2023; Osvath & Gärdenfors, 2005; Snyder et al., 2022; Tennie et al., 2017; Vaesen, 2012; Völter & Call, 2014), and changes in brain size throughout hominin evolution (Aiello & Wheeler, 1995; Carroll, 2003; D. Lieberman, 2011; Burini & Leonard, 2018; Ponce de Leon et al., 2021; Zollikofer et al., 2022).<sup>2</sup> In the following sections, we connect a variety of changes to the morphology of the articulatory complex throughout human evolution in the context of inferred changes to hominin behavior including cooking, tool use and manufacture, social behavior, and morphology. Reflecting the relative sparsity of the early Pliocene fossil record, we focus on paleoanthropological evidence from the late Pliocene and Pleistocene epoch, but where available also draw upon relevant data from extant great apes. We conclude with speculations about the origin of syllabic vocal production in the hominin lineage.

## Evolution of the Vocal Tract

The vocal tract was significantly reconfigured in hominin evolution, with non-human great apes exhibiting a different shape and position of the hyoid bone (Falk, 1975; Steele et al., 2013), and a short and narrow pharynx and expansive oral cavity (Bermejo-Fenoll et al., 2019; Negus, 1949; Sato et al., 2023), compared to those of anatomically modern *H. sapiens*, or modern humans. Anatomical reconfiguration of the vocal tract, involving expansion of the pharynx, permanent descent of the tongue root and larynx into the throat, and rounding of the tongue body has been widely regarded as an adaptation for speech (Negus, 1949; Lenneberg, 1967; P. Lieberman, 1984, 2012, 2017; Fitch, 2000; de Boer & Fitch, 2010; D. Lieberman, 2011; Ekström & Edlund, 2023a; Sato et al., 2023). Some animals like big cats also have low larynges (Weissengruber et al., 2002), but their tongues remain anchored in the oral cavity, reflecting disparate evolutionary pressures. As summarized by P. Lieberman (2006, p. 278), “A low larynx does not signify an SVT that can produce the full range of human speech”.

The derived form of the human SVT is achieved during postnatal ontogeny and human infants are born with vocal tracts resembling those of non-human primates, with a high larynx, narrow pharynx, and tongue contained in the oral cavity. Throughout early childhood, the mouth of human infants is shortened, in relative terms, through a rotation of the skeletal structure supporting the palate (D.

<sup>2</sup> Readers of this text may be familiar with the broader literature on the evolution of language, where a prominent debate concerns the ostensible “manual”, “vocal”, or “multimodal” origins of language. In the final section, we speculate on potential applications for this broader literature. However, we emphasize that this paper is not primarily concerned with the evolution of language or speech – but with the emergence of the human articulators themselves and their apparent correlates in hominin evolution.

Lieberman et al., 2000), gradual descent of the larynx and tongue root (D. Lieberman & R. McCarthy, 1999), and lengthening of the neck (Mahajan & Bharucha, 1994), ultimately achieving the roughly equally-proportioned horizontal (oral cavity and oropharynx) and vertical (pharynx) sections of the vocal tract that characterize the adult configuration (D. Lieberman & R. McCarthy, 1999; Vorperian et al., 2005, 2009; Moran et al., 2024). This configuration is dangerous in the sense that it results in an increased risk of choking during swallowing as a bolus of food passes over the laryngeal opening, which is covered by the epiglottis but not sealed off by a locked soft palate-epiglottis, the configuration in other animals. Choking on food remains a cause of death in modern human; as such, the reorganization of the hominin vocal tract would appear to reduce reproductive fitness (Palmer et al., 1992). Positive selection pressure must have outweighed negative selection for choking.

The first attempt at determining the phonetic capacities of non-human primate vocal tracts was undertaken by P. Lieberman et al., (1969, 1972), who investigated the phonetic capacities of a rhesus macaque (*Macaca mulatta*). The apparent inability of non-human SVTs to articulate “quantal” vowels (Stevens, 1969, 1972, 1989) formed a cornerstone of Lieberman’s theory of spoken language evolution (P. Ekström, 2024; Lieberman, 1984, 2012). In this view, non-human animal vocal tracts lack the capacity to produce the full extent of human speech sounds, including vowels [a] (“ma”), [i] (“see”), and [u] “true” [u] (“boot”) *in the same way as humans* – sounds found nearly universally across human spoken languages (Moran & McCloy, 2019) and held by Lieberman (1984) to be noteworthy for their articulatory and perceptual distinctiveness (Peterson & Barney, 1952; Nearey, 1978; Stevens, 1989; P. Lieberman, 1984, 2012, 2017; Friedrichs et al., 2017; Friedrichs & Dellwo, 2023; but see Diehl, 2008). Recent data has nuanced this account, though not fully refuted it. A study by Fitch et al. (2016) replicated Lieberman’s macaque study, confirming the “Lieberman account” (P. Ekström, 2024; Lieberman, 2017) and showing that macaques cannot produce the full extent of human vowel space—even allowing for extreme contortions of the mandible and pharyngeal muscles involved in yawning (Ekström, 2024; Everett, 2017). Namely, their capacities do not include the full extent of the human vowel space.

Further, while [u]-like sounds may be approximated by other animals, including baboons (Boë et al., 2017), chimpanzees (Grawunder et al., 2022), and orangutans (Ekström et al., 2023), they cannot be reproduced identical tongue gestures as those employed by modern human speakers (P. Berthommier, 2020; Berthommier et al., 2017; de Boer & Fitch, 2010; Ekström, 2024; Lieberman et al., 1972; Nishimura, 2005; Takemoto, 2008). Boë et al. ascribed human articulation of [u] and [a] to baboons based on analogy to human speech data without reference to *in-situ* articulatory data from baboons, neglecting the possibility that such vowel-like properties likely result from species-unique constraints. Reflecting the comparative shape of human and baboon tongues and SVTs (Berthommier et al., 2017; Negus, 1949), such articulation is impossible. Rather, alternative gestures appear to explain these vowel-like properties (Berthommier, 2020).

Thus, morphology apparently prevents humanlike articulation involving deformation of the tongue body to the extent required for these vowel sounds, and recent modeling efforts suggest that the formant dispersions estimated from chimpanzee

“hoo’s” may be achieved through other articulatory configurations (Ekström & Edlund, 2023b). Grawunder et al. (2022) document the occurrence of [a]-like formants in chimpanzee “barks,” but these vocalizations are uttered with a distinctly lowered mandible beyond the range required for “true” [a], a situation un conducive to fluid coarticulated speech. In short, the occurrence of “vowel-like” formant dispersions does not necessitate (or even imply) that such utterances are produced in the same way as their seeming human language counterparts. Rather, the extreme mandible positions employed indicate that articulatory configurations are more costly in comparison.

## Late Pliocene

### Why the Long Face? The Early Hominin Vocal Tract

The nature of the hominin fossil record, in particular the lack of preservation of soft tissues including the tongue and other elements of the SVT, means that there are only hints regarding the phonetic capabilities of our fossil ancestors and close relatives (Lieberman & R. McCarthy, 2015; Clark & Henneberg, 2017). By the appearance of *Australopithecus afarensis* ~3.7 – 3.0 million years ago, early hominins were obligate bipeds. The transition to upright walking may also have facilitated the evolution of fine breathing control using the thoracic muscles at some unknown point after 1.6 million years ago, as inferred by differences in vertebral canal proportions between *Homo erectus* and *H. sapiens* (Hewitt et al., 2002; MacLarnon, 1987; MacLarnon & Hewitt, 2004). Once the head was no longer tethered to the thorax (Bramble & D. Lieberman, 2004) the hominin neck could more freely vary in relation to ecogeographic parameters; in the words of Bramble and Lieberman (2004, p. 350), “cranially oriented glenoid cavities (present in *Australopithecus*) ... would tend to ... minimize axial rotation of the head” (see also Sato et al., 2023). There is incremental evidence that australopiths possessed laryngeal air sacs (Alemseged et al., 2006), a configuration that has been argued to impede speech (de Boer, 2012). However, in the “articulator-call” acoustical scheme noted by Grawunder et al. (2022), chimpanzees are shown to produce [a]-like and [u]-like extremes delimiting their vowel-like space in continuity with human speech, suggesting that air sacs could have a limited effect on speech production.

Following works by Laitman (Laitman & Heimbuch, 1982; Laitman et al., 1979), Crelin (1987, 1989) performed a series of investigations on cranial and speech-centric morphology, arguing that the skulls of australopiths and *Homo habilis* were – with regards to speech capacities – essentially “apelike”, whereas *H. erectus* skulls were intermediate in form between the earlier australopiths and *H. sapiens*. According to this view, the basicranial prerequisites for speech arose late during hominin evolution, and only recent hominin lineages would have evolved the capacity to produce the full range of human speech sounds. Crelin’s interpretations with regards to the basicranium are challenged by more recent works (Gunz et al., 2020; Ponce de Leon et al., 2021) that are agnostic as to the relationship between basicranial anatomy and vocal tract proportions. Crelin (1987) argued that the full extent of

modern human speech capacities likely had evolved recently in human evolution. Later developments revealed a number of problematic assumptions (discussed in Sect. "Neanderthal speech and the late origins of the modern human vocal tract") implemented in the Crelin reconstructions, ultimately rendering these efforts ambiguous. Nonetheless, there is ample evidence of substantial evolution of several speech articulators throughout hominin evolution, most prominently involving the jaw and other craniofacial features.

During the course of evolution, the face of *H. sapiens* has undergone a rapid reduction (D. Katz et al., 2017; Lieberman, 2011; Lordkipanidze et al., 2013; von Cramon-Taubadel, 2017; Zollikofer et al., 2022). While many non-human mammals are prognathic, with faces protruding anterior to the anterior cranial fossa and the frontal lobes of the brain, the modern human face is almost completely orthognathic (flat) and the anterior-most end of the vocal tract (i.e., the lips) protrudes only marginally anterior to the anterior cranial fossa. This is significant for two reasons. First, reduction in the size of the oral cavity and oropharynx roughly equalizes the lengths of the horizontal and vertical segments of the vocal tract. Second, prognathic animals achieve variable vowel qualities by alternately “flaring” and elongating their vocal tracts, movements accomplished by lowering and raising the mandible (Shipley et al., 1991; Schön Ybarra, 1995; P. Lieberman, 2012; Schötz, 2020; Goncharova et al., 2024; Ekström et al., 2024b). As such, the loss of prognathism meant an effective loss of such “coasting effects” of elongated horizontal vocal tracts.

Retraction of the face below the frontal lobes of the brain, flexion of the basicranium, and shortening of the face and underlying naso- and oropharynx (D. Lieberman et al., 2002; Trinkaus, 2003) create a “spatial packing problem” on the underside of the cranium, eventually necessitating laryngeal descent, untethering the hyoid from the lower border of the mandible. Importantly, the oral cavity of non-human great apes, per se, likely does not impose limits on possible speech production; it is interaction with other elements that create meaningful pressures on any speech behavior. The shortening of the oral cavity, descent of the tongue root into the throat (along with permanent descent of the larynx), and expansion of the pharyngeal cavity effectively unlocks the extremes of phonetic potential exploited today universally by human speakers (P. Lieberman et al., 1972, 1992; Laitman, 1983; Carré et al., 1995, 2017; de Boer, 2010; P. Lieberman, 2012). The concomitant rounding of the tongue also makes possible the fine distinction between various articulatory targets (Ekström & Edlund, 2023a; Gay, 1974; Lindblom, 1963; MacNeilage, 1998; Öhman, 1967; Studdert-Kennedy, 1998) unavailable to hypothetical speakers equipped with “unconfigured” vocal tracts.

Movements of the mandible, particularly the oscillatory actions of opening and closing, influence the amplitude modulation of the speech signal, which in turn shapes the syllabic organization of speech. Consequently, the mandible has often been regarded as a “serial organizer” of speech patterns (Barlow & Estep, 2006; Gracco & Abbs, 1988; Lund & Kolta, 2006; MacNeilage, 1998). In everyday speech, humans typically produce about four syllables and around 12 phonemes per second (Levelt, 1999; Poeppel & Assaneo, 2020), though even faster rates are attainable. Interesting parallels have been observed in a variety of non-human primates (Bergman, 2013; Ghazanfar and Takahashi, 2014) including great apes (Lameira et al.,

2015; Pereira et al., 2020) where lip movements and smacks, resembling “speech-like rhythms” within the frequency range of 3–8 Hz, have been documented. These findings suggest that the fundamental biological basis of syllabic rhythms might share a common ancestry. A recent study by Piette et al. (2022) found that vocalization patterns across 89 masticating species predominantly manifest within this frequency range. As properties of the mandible differed vastly between species, such patterning is suggestive of broad biological mechanisms at play influencing these rhythms, beyond the specific nuances of mandible morphology. Yet, this overarching biological rhythm does not fully encapsulate its role in the temporal organization of human speech. Recent work by Friedrichs and Dellwo (2022) posits that longer mandibles in modern humans can potentially limit syllable production rates, especially under conditions necessitating rapid articulation. This might suggest that, even within the broad biological rhythm observed, the specifics of mandible morphology may impose a cap on possible syllable production rates, delineating the upper boundaries of syllable repetition. Phonetic consequences of mandibular morphology, thus, should not be overlooked.

Mandibles of extant non-human great apes have a simian shelf – a boney horizontal ridge projecting inward from the inside of the mandible, effectively thickening the bone at the lower border of the mandibular symphysis. This shelf is somewhat reduced in early hominins, but a postincisive plane is nonetheless meaningfully expressed in australopiths and early *Homo*, including *H. erectus* (e.g., Strait & Grine, 2004). P. Lieberman and colleagues (1972) argued that the simian shelf would preclude proper articulation of “true” back rounded vowel [u], which involves the creation of a narrow pocket in front of the mandibular incisor teeth. Relevant modeling suggests that apparently vowel-like vocalizations by non-human primates are produced differently than they are in modern humans (Berthommier, 2020), consistent with this idea. Australopiths possessed multiple cranial adaptations for masticating mechanically resistant foods (e.g., Jolly, 1987; Peters, 1987; Strait et al., 2009, 2013; A. L. Smith et al., 2015a, 2015b), including larger mandibles (Demes & Creel, 1988; Humphrey et al., 1999). The later reduction of this feature in *Homo* is widely considered to signify some type of shift in diet or food processing. Consistent with this idea, Stedman et al. (2004) have argued that the gene encoding the predominant myosin heavy chain expressed in chimpanzee masticatory muscles was inactivated in the lineage leading to *Homo* around ~2.4 mya, a change associated with reductions in size of both individual muscle fibres and the total size of masticatory muscles.

The mandibles and teeth together constitute a primary weapon for extant non-human great apes (including chimpanzees and gorillas) when engaging in confrontations and combat with conspecifics (Hill et al., 2001; Kortüm & Heinze, 2013), and in chimpanzees when hunting (Goodall, 1986; Wrangham, 1975). Reduction of the canine teeth in early hominins may denote the loss of this functional role for the mandible in australopiths and later hominins. The loss of “ape-like” mandibular robustness may have been possible in early hominins for all the reasons listed above, e.g., because much of food processing was already outsourced to the hands and tools (Schick & Toth, 1994; Toth & Schick, 2018). In a hypothetical hominin lacking behavior such as consistent use of stone tool technology, the necessity of greater

masticatory forces may have prevented any loss of mandibular robustness from taking hold in earlier hominin populations. The gradual reduction of the mandible likely eased limitations on would-be syllable production rate, enforced on human ancestors and collateral relatives through the comparative robustness of their mandibles. Insofar as a gracile mandible is a pre-adaptation of modern human syllabic speech, it is reasonable to infer that changes in food processing effectively served as a pre-adaptation event (or series of events) that ultimately set the stage for its evolution.

## Early Tools of the Trade

Archaeological records suggest that while early hominins likely engaged in mechanical processing of the kind commonly found in extant non-human great apes, the invention of stone tool technology may have marked the beginning of additional cognitive evolution. Extant chimpanzees appear to plan for the future (Bräuer & Call, 2015; Mulcahy & Call, 2006; Osvath, 2009; Osvath & Osvath, 2008) and certainly make use of tools, including for food acquisition (Boesch & Boesch, 1990; Johnson-Frey, 2003), though important distinctions may yet be made with regard to such behaviors. In the words of Van Casteren et al. (2022), “for our ancestors, before the onset of cooking and sophisticated food processing methods, the costs [of chewing] must have been relatively high” (see also Schick & Toth, 1994). Because neither humans nor chimpanzees have obvious signs of dental adaptations for chewing meat specifically, Wrangham and Conklin-Brittain (2003) argue that early hominins made systematic use of meat tenderizing techniques. Even rudimentary slicing and pounding techniques would have drastic effects on chewing time, with Zink and D. Lieberman (2016) concluding that “selection for smaller masticatory features in *Homo* would have been initially made possible by the combination of using stone tools and eating meat.” This is suggestive of a meaningful relationship between the would-be speech articulators and the ecological constraints which may have limited their “pre-adaptiveness” to speech evolution. In addition, softening food via putrefaction and/or fermentation (Speth, 2017) could have facilitated the same changes in the chewing apparatus. These approaches are, however, significantly more difficult to track than stone tools, which is why we shall focus on such tools in this account.

Simple, putative stone tools with sharp cutting edges, as well as cut-marked bones, appear in the archaeological record approximately 3.4 – 3.3 mya in East African landscapes occupied by early hominins such as *A. afarensis* and *Kenyanthropus platyops* (Harmand et al., 2015). Oldowan stone tools – consisting of simple choppers, flakes, and spheroids, the earliest stone tools to be widely accepted as such in the current literature – have been associated with various early pre-modern hominins. These tools first appear between 3.0 and 2.5 million years ago (Semaw et al., 2003; Plummer et al., 2023) and subsequently become increasingly common in the African archaeological record. It is impossible to know for certain which hominins made these tools, but representatives of *Australopithecus*, *Paranthropus* and *Homo*



existed in Africa at the relevant times (Harmand et al., 2015; Heinzelin et al., 1999; McPherron et al., 2010).

The use of such tools for food processing may have been varied, from the cracking of nuts and hunting – as have been observed, e.g., in wild chimpanzees (Boesch & Boesch, 1983; Goodall, 1986) – to the cracking of animal bones and butchery (Boëda et al., 1999; Jacob-Friesen, 1956; Keeley, 1980). Osvath and Gärdenfors (2005) have argued that the origins of “anticipatory cognition” – the ability to mentally represent future needs – is reflected in Oldowan stone tools (see also Toth & Schick, 2018), although even earlier hominins occasionally developed similar stone technologies (Harmand et al., 2015; Lewis & Harmand, 2016; Panger et al., 2002; Semaw et al., 1997, 2003). The invention of Oldowan tools does not necessitate cultural copying of know-how (Snyder et al., 2022), suggesting that later (and perhaps much later) tool-making capacities may have been maintained and refined through cultural transmission proper.

### A Brain Made for Speaking?

Endocranial volume has increased fourfold in hominins over the past two million years (D. Lieberman, 2011; Hawks, 2011; Montgomery, 2018; DeSilva et al., 2021). Estimates for chimpanzee endocranial volume ranges from 282 cm<sup>3</sup> to up to 557 cm<sup>3</sup> (Herndon et al., 1999; Isler et al., 2008; Neubauer et al., 2012; Tobias, 1971; Zihlman et al., 2008). In comparison to the engineering constraints on non-human primate vocal tracts with regard to producing the full range of human speech sounds, the capacity for mapping articulatory targets may be a product of neural evolution, as evident from the modest success of non-human great apes subjected to rigorous speech (Ekström, 2023; Hayes & Hayes, 1951) and sign language exercises (Gardner et al., 1989). Apes are apparently limited with regards to learning many new vocal behaviors, even when subjected to human tutorship (Ekström, 2023; Ekström et al., 2024a; Lameira, 2017; Shofner, 2024). Australopiths may have evolved neural but not peripheral speech substrates. Compared to chimpanzees, the australopith brain was likely slightly larger, and there are signs that it is reorganized in *A. africanus* (Holloway et al., 2004; Tobias, 1968, 1971), *A. sediba* (Carlson et al., 2011) and *A. afarensis* (Gunz et al., 2020).

Unfortunately, while recent advances have allowed researchers to infer something of the presence and extent of Broca’s and Wernicke’s areas (Brodmann’s areas 44 and 45, and 22 respectively) from hominin crania (Hill & Beaudet, 2023), two main impositions pose problems for inferring their relevance to speech evolution. First, the exact contributions of increasing brain size to speech is unknown. More significantly, however, a growing body of neurolinguistics literature has deemphasized the contributions of Broca’s and Wernicke’s areas to speech, pointing instead to a crucial role for subcortical circuitry in speech production (Alexander et al., 1987; Alm, 2021; Dronkers et al., 2007; Ekström, 2022; Guenther, 2016; Hodgson & Hudson, 2018; Lashley, 1930, 1951; Lieberman, 2000; Murdoch, 2001; Pidoux et al., 2018). For example, input from the cerebellum is thought to regulate speech production, facilitating the temporal organization of speech into rhythmic utterances

(Ackermann, 2008). As no trace of subcortical neurons are left in fossilized crania, such insights may be forever beyond recovery. That being said, across species, there is a strong positive correlation between sociality and (relative) brain size (Connor, 2007; Dunbar & Schultz, 2007; but see Lindenfors et al., 2021), and some have argued that primates, living in large social groups where individuals have to keep track of the identities and interactions of individuals and their kin, provided particularly fertile ground for future such social evolution (Seyfarth & Cheney, 2014; van Horik & Emery, 2011).

## The Pleistocene

### Tools and Cooking

Cooking has been argued to be a “biological trait” in modern humans (Wrangham & Conklin-Brittain, 2003). With exceptions (McCauley et al., 2020), modern hunter-gatherers are generally well-versed in creating and using fire (Gott, 2002), and all have been known to eat cooked food (Wrangham & Conklin-Brittain, 2003). Even extant non-human great apes prefer cooked food (Wobber et al., 2008), and chimpanzees have been shown to delay consumption of raw food in exchange for its cooked equivalent later in time (Warneken & Rosati, 2015). Such behavioral ubiquity across hominids suggests substantial benefits from cooked foods. Indeed, a variety of findings are suggestive of such benefits.

First, food processing by fire may serve to purify even meat scavenged from other sources; many tribes of modern hunter-gatherers obtain a substantial portion of their consumed meat from scavenging (Yellen, 1991; O’Connell et al., 1998; but see also Lupo & Schmitt, 2005). It has been argued that animal foods are a necessity of modern human diets, prior to the invention of agriculture (Larsen, 2003) some ~12 kya, judging by the relative caloric poverty of other available food stuff. Bone marrow, subject to less extensive bacterial growth compared with meat (A. R. Smith et al., 2015a, 2015b; see also Speth, 2017), may have provided early scavenging hominins with a source of food, even when hunting was not an option. Hominins were likely consuming meat and bone marrow as early as 2.5 mya (Blumenschine & Pobiner, 2007; Cáceres et al., 2023; Plummer et al., 2023) and more certainly by 1.9 mya (Pante, 2013; Pante et al., 2018), but occasional meat consumption likely goes back further still (McPherron et al., 2010)—not least given that apes are known to consume meat. Thus, for many scavenging species, possibly including early *Homo*, cooking may have provided a way to purify meat, which may have otherwise accumulated dangerous bacterial loads (Ragir et al., 2000; A. R. Smith et al., 2015a, 2015b; Speth, 2017). Scavenged meat may thus have provided a valid, substantial source of nutrients, even in times where active hunting was not possible. A combination of scavenging and small-game hunting may have been a likely starting point in the evolution of hunting strategies involving larger game.

Cooking may result in a substantial increase in the caloric density of foods (Wrangham, 2011; Laird et al., 2016; but see also Zink & D. Lieberman, 2016; Cornélio et al., 2016). Accordingly, living humans on raw food diets (and who thus



never or rarely consume processed foods) have been reported as exhibiting lower-than-average body weight, and body mass index has been found to be inversely correlated with the proportion of raw food in the diet (Carmody & Wrangham, 2009; Koebnick et al., 1999). Lombard and van Aardt (2023) argued against this hypothesis, however. By analyzing the rich variety of plant foods available in the Klasies River area in South Africa, they found that a majority of species in the catchment area can be consumed raw. Thus, it may very well be possible to survive given such a diet (presumably including raw meat). Moreover, putrefaction and fermentation (neither of which requires extensive know-how) can further increase digestibility (Speth, 2017). Finally, cooking typically softens food (Wrangham et al., 2009), decreases digestion time, and increases digestibility of many different starches (Kataria & Chauhan, 1988; Sagum & Arcot, 2000; Smith et al., 2001) and plant proteins (Chitra et al., 1996).

Chimpanzees spend up to half their time awake each day masticating food stuff (Ross et al., 2009; Wrangham, 1977), while modern humans spend only around 4.7% of their daily activity doing the same (Organ et al., 2011; Zink & D. Lieberman, 2016). Based on observations of chimpanzees in the wild (Goodall, 1986; Wrangham, 1975), Wrangham and Conklin-Brittain (2003) have estimated chimpanzee caloric intake from meat at around 400 cal per hour. At the same rate, modern humans would need to spend ~6 h per day simply satisfying energy needs (see also Van Casteren et al., 2022). To break down hard foods, modern humans cook, ferment, putrefy, cut, slice, pound, blend, combine, chemically alter, and artificially breed various food sources. The multiple ways by which modern humans manipulate and process their food, and the timing of emergence of these behaviors, have accordingly become the focus of significant recent research.

In comparison to mechanical processing, however, the deliberate application of heat to food is found exclusively in *Homo* (Wrangham, 2009; but see Warneken & Rosati, 2015; Jacobs et al., 2021). It has been suggested that the earliest impact of cooking in the known hominin fossil record is associated with *Homo erectus* (1.9 mya), who exhibited significantly reduced teeth and mandibles compared to earlier hominins (Wrangham et al., 1999; Wrangham, 2007; D. Lieberman, 2011). This estimate is controversial, however, and, given the evidence reviewed above, unlikely. There was also a reduction of mandibular muscle myosin, oral cavity volume (Lucas et al., 2006), gut volume (particularly the cecum and colon), which resulted in faster gut passage rates (Milton, 1987; Aiello & Wheeler, 1995; Chivers & Hladik, 1984; Martin et al., 1985; Hladik et al., 1999; see also Ben-Dor et al., 2011; Zink et al., 2014; Zink & D. Lieberman, 2016; Wrangham, 2017), shorter faces relative to body size (D. Lieberman, 2011), and a substantial increase in brain size (Aiello & Wheeler, 1995; Carroll, 2003; D. Lieberman, 2011; Burini & Leonard, 2018; Clark & Henneberg, 2021, 2022). Parsimoniously, all of these changes may be related to the increasingly widespread use of techniques that eased ingestion and/or digestion. For example, the use of stone tools (for pounding, cutting, slicing, etc.), putrefaction, fermentation, and even retrieving naturally cooked food in the aftermath of natural fires (“fire foraging”), would to various degrees have facilitated observed anatomical changes. In a later section, we explore the impact of these developments on possible speech production capacities.

Food processing, broadly considered, includes both mechanical processing (the manipulation of would-be food stuff via cutting, slicing, etc.) and cooking – the application of heat to food (Wrangham, 2009; D. Lieberman, 2011; Zink et al., 2014; Zink & D. Lieberman, 2016; Gowlett, 2016; Speth, 2017). All such processing may be conceived as preprocessing food outside the body itself, with the benefits of increasing caloric density, decreasing time spent chewing and digesting, and consequent reduction of toxins and parasites (Stahl et al., 1984; Ragir, 2000; Wrangham & Conklin-Brittain, 2003). Mechanical processing is readily observable across extant non-human primates, and thus such manual processing also preceded cooking in evolution. While prosimians such as lemurs make extensive use of the tongue for food manipulation (Iwasaki et al., 2019), the evolution of opposable thumbs in early primates meant a functional transition toward manual dexterity and manipulation of food stuff actively using the hands and fingers (Dew, 2005; Heldstab et al., 2016, 2020; Pal et al., 2018; Tan et al., 2016). Among primates, humans have especially dexterous hands (with broad, fleshy fingertips and proportionally long thumbs whose muscles allow movement independent of the other digits) enabling an enhanced precision grip, a configuration that likely emerged early in the *Homo* lineage (Diogo et al., 2012; Karakostis et al., 2018, 2021; Leakey et al., 1964; Marzke & Shackley, 1986; Marzke, 1997, 2013; Susman, 1994, 1998). However, manual food processing is also evident in extant great apes, such as orangutans opening fruit (Stoinski & Whiten, 2003) and chimpanzee termite-fishing (Sanz et al., 2009). The complexity of such processing techniques pales in comparison with that of Pleistocene hominins, however. Specifically, at some point after 2.0 million years ago, and certainly by 1.7 million years ago, more refined Acheulian tools in the form of hand-axes and large bi-facial cleavers appear in the African archaeological record. It is not possible to ascertain what populations created these tools, but their production is generally attributed to *H. erectus* (Jurmain et al., 2005; Keeley, 1980; Rose & Marshall, 1996) before the transition into the Acheulean technological niche (overview in Toth & Schick, 2018).

While the timing of the origin of cooking is subject to great dispute, it is typically assumed to coincide with control of fire (implying fire *making*). Archaeologically-based estimates for this event based on earth-ovens found at various archeological sites across Europe and the Middle East date from ~250 kya (Brace, 1999; James et al., 1989), 400 kya (Barkai et al., 2017; Roebroeks & Villa, 2011; Shimelmitz et al., 2014), and 780 kya (Goldberg et al., 2023; Walker et al., 2020; Zohar et al., 2022), whereas estimates based on the appearance of *H. erectus* range as early as 1.9 mya (Wrangham, 2009; Wrangham et al., 1999). More recent dates are archaeologically well-supported, while earlier estimates are based on rare (Roebroeks & Villa, 2011) and/or localized phenomena (Shimelmitz et al., 2014; Zohar et al., 2022).

Others have suggested, partly based on widespread anatomical differences present already in early-Pleistocene *Homo*, place the control of fire around 1 mya (Berna et al., 2012; Fernández-Jalvo et al., 2018) or even 1.5 mya (Hlubik et al., 2019; see also Wrangham, 2017). However, it is difficult to exclude natural fires as potential sources (e.g., Roebroeks & Villa, 2011). The earliest use of fire may not have necessitated control of the element and may be invisible to archaeological

inquiry. Early *Homo* may have made use of accidental or natural fires (Gowlett & Wrangham, 2013) and even extant chimpanzees live and forage in the presence of recurring natural fires (Pruetz & Herzog, 2017; Pruetz & LaDuke, 2010). Finally, it has been argued that Neanderthals may have lived for generations without fire in times and places absent frequent natural fire (Abdolazhadeh et al., 2022). Overall, thus, cooking likely played a major role in human evolution, but may have done so only at a relatively late date (cf. Roebroeks & Villa, 2011; Shimelmitz et al., 2014).

## Emergent Social Consciousness and Culture

One particularly salient marker of organization of primate social systems is sexual dimorphism, the phenomenon in which males and females of the same species exhibit different sizes and/or shapes. Effectively, across non-human primates, where there is strong competition for mating access to females, male dimorphic traits are exaggerated (Plavcan, 2001). In the polygynous *Pongo* and *Gorilla* genera, males may be more than twice the size of females, while in monogamous gibbons (with reduced male-male competition), the two sexes are essentially monomorphic. Chimpanzees, living in multimale-multifemale societies, exhibit a male–female body mass ratio of ~1.3:1, while modern humans exhibit a ratio of ~1.15:1 (Dixon, 2008). The evolution of this relationship in ancestral hominins, however, is not straightforward.

While there is broad agreement that australopiths exhibited strong body size dimorphism beyond the levels seen in extant chimpanzees, the australopith condition is atypical. Normally, across primates, the relative size of canine teeth constitutes a strong marker of sexual dimorphism in a species (Plavcan and Schaik, 1992; Lee, 2005). However, australopiths likely exhibited a highly “unusual combination” (Lockwood, 1999, p. 98) of sexually dimorphic features characterized by low canine dimorphism and high body dimorphism, while both australopiths and early *Homo* exhibited significant male–female size dimorphism (Plavcan & van Schaik, 1997). Due to ecological changes, mid-Pleistocene *Homo* may have collaborated in food quests, rather than foraging individually (see Tomasello et al., 2012; Wrangham, 2019). In this hypothetical scenario, more-collaborative, less-selfish individuals may have been favored by natural selection. To this day, chimpanzees occasionally engage in hunting and meat consumption (Goodall, 1986; Wrangham, 1975), and while seemingly less common, similar observations have also been reported in bonobos (*P. paniscus*) (Hohmann & Fruth, 2007). Chimpanzee hunting is mostly selfish, however (Tennie et al., 2009), and the timing of the origin of habitual social hunting and meat consumption is uncertain.

Wrangham (2019) has argued that a process of self-domestication, including reduced aggression, an emergence of cooperative breeding, and increased self-control, likely placed selection pressure on anatomical features, both for signaling behavior and means of communication (see also Cieri et al., 2014; Thomas & Kirby, 2018; Benítez-Burraco & Kempe, 2018; but see Sánchez & van Schaik, 2019). Anatomical changes such as a narrowing of the male face have been taken as consistent with a selection pressure favoring reduced reactive aggression during the last

300,000 ky in *H. sapiens* (Gärdenfors et al., 2012; Leach, 2003; Wrangham, 2019). This time window overlaps with the estimate proposed by the McCarthy reconstructions indicating the emergence of modern human vocal tract dimensions in early *H. sapiens* (P. Lieberman & R. McCarthy, 2015) and follows immediately upon the proposed time of mutation of the Forkhead box protein P2 (FOXP2) gene by Krause et al. (2007), believed meaningful for linguistic evolution in early humans, and shared with Neanderthals, some 300–400 kya.<sup>3</sup>

This estimate overlaps with the proposed time span earmarked for the emergence of cumulative culture of know-how by Tennie (2023), sometime around ~500 kya. Tennie (2023) places the transition cultural know-how relatively late in hominin evolution, sometime around ~500 kya. For potential phonetic systems, this may be meaningful; while great apes have the capacity, however exceptionally, to learn novel vocalizations (Ekström et al., 2024a; Lameira et al., 2016; Russell et al., 2013; Salmi et al., 2022a; Wich et al., 2009), there is to date no strong evidence that such forms are preserved and transmitted across generations. Yet every new human infant acquires and effectively replicates the phonetic systems of their caretakers (Thelen, 1991; Vihman, 2014). As such, the potential relevance of cultural transmission in the development and maintenance of phonetic systems is an intriguing avenue for future work.

## Neurocognitive Adaptations

A notable shift in endocranial volume took place from between ~500 cc and ~750 cc in *H. habilis* and similar early *Homo* species to between ~900 cc and ~1150 cc in all but the earliest populations of *H. erectus* (Cornélio et al., 2016; Leakey, 1966; Leigh, 1992; McHenry, 1992; Pu et al., 1977). Estimates for specimens from Dmanisi, which many researchers consider to represent early *H. erectus*, possessed endocranial volumes between ~550–775 cm<sup>3</sup>—closer to values for *H. habilis* (Rightmire, 2013; Vekua et al., 2002). Another increase is observed between *H. erectus* and *H. sapiens*, with *H. sapiens* (Holloway, 1997; White et al., 2003) exhibiting an endocranial volume of ~1350 cm<sup>3</sup>. A modern interpretation of the paleo-anthropological and archaeological records suggest that neural organization – not volume alone – need be taken into account when considering cognitive evolution (Du et al., 2018; Logan et al., 2018; Montgomery, 2013, 2018; Ponce de León et al., 2021; Shultz et al., 2012). Modeling efforts aimed at elucidating the temporal trajectory of hominin brain evolution by DeSilva et al. (2021) estimated a changepoint

<sup>3</sup> To date, the oldest available hominin genetic data are from Middle Pleistocene hominins, approx. 400 ky old (Meyer et al., 2016). As such the mutation of the gene likely followed the evolution of most major vocal tract morphological adaptations discussed in our account. In addition, we reiterate that our work is only tangentially related to language evolution per se. Involvement of FOXP2 in speech may involve facilitating general planning of speech-related motor movements (Morgan et al., 2023). However, its expression has not generally been associated with morphological elements involved in speech production. As such, we limit our inclusion of this literature to the date of the gene's mutation as generally argued in current literature, and as a potential adaptation (“correlate”) and factor facilitating the emergence – or modification – of speech production capacities in the Upper Pleistocene.

at  $2.10 \pm 0.07$  mya, around the first appearance of *H. erectus* in the fossil record (Herries et al., 2020). This changepoint likely reflects the beginning of a trend that appears first in early *Homo* populations and then persists and becomes more evident in later populations of *H. erectus*.

This marked increase in brain size has alternately been ascribed to increased consumption of meat (Bunn, 2007; Leonard et al., 2007; Milton, 1999; Speth, 2017), tubers (Wrangham et al., 1999), and aquatic foods (Broadhurst et al., 2002) in early hominins. *H. habilis* and *H. erectus* may also have possessed lateralized frontal lobes (Cantalupo & Hopkins, 2001; Holloway et al., 2004) and larger brains in comparison to earlier hominins (Holloway et al., 2004). Yet other accounts emphasize sociality, arguing that increased social demands effectively put a premium on cognitive skills to navigate those demands (increasing brain size in the process) (Dunbar, 1998; Isler & van Schaik, 2012). Anthropological works show that consumption of cooked foods in modern hunter-gatherers – like hunting itself – is often communal (Hayden, 2014; Whitehead, 2000; Jones, 2007; Dunbar, 2014), blurring the lines between accounts: access to beneficial foods may have required collaborative hunting. Genetic evidence also shows that the FOXP2, believed meaningful for language (Enard et al., 2002; Fisher & Scharff, 2009; Zeberg et al., 2024), was present in Neanderthals (Krause et al., 2007).

### Neanderthal Speech and the Late Origins of the Modern Human Vocal Tract

Contrary to popular sentiment, it has long been a consensus among researchers that Neanderthals likely possessed a form of language (P. Johansson, 2015; Lieberman et al., 1972). Pioneering research on the question of hominin phonetic capacities was conducted by Lieberman and colleagues (P. Lieberman & Crelin, 1971; P. Lieberman et al., 1972). Their work involved the reconstruction of the SVT of the La Chapelle-aux-Saints Neanderthal skull, whose phonetic capacities were inferred by use of a computer program to explore all possible vocal tract configurations that could be fit to the reconstructed basicranium and neck (Henke, 1966). From the resulting vowel space (the two-dimensional area denoting  $F_1$ - $F_2$  correlations of vowels, with extremities corresponding to articulatory extremes), the authors observed that – like predictions for vocal tracts of non-human primates (P. Lieberman et al., 1969, 1972) – the Neanderthal vowel space did not include the full extent of human vowels. The reason for this apparent limitation was Neanderthals' relatively longer faces and shorter necks and pharynges, precluding fully modern human-like production. Specifically, the uniquely acoustically sensitive regions of the upper and lower pharynx could not be sufficiently navigated (Carré et al., 2017; Stevens, 1989; Wood, 1979, 1986). Boë et al. (1999) claimed that the conclusions of the Lieberman-Crelin efforts were that Neanderthals had been incapable of speech, but this is a misunderstanding (Ekström, 2024). In reality, Lieberman and colleagues (1972, p. 303) concluded that it was "... likely that Neanderthal man's linguistic abilities were at best suited to communication at slow rates and at worst markedly inferior at the syntactic and semantic levels to modern man's linguistic ability. Neanderthal man's language is an intermediate stage in the evolution of language." According to

this view, a modern human vocal tract was necessary for the extent of fully modern human speech; Neanderthal speech, limited by a relatively short neck, would have been possible, but less efficient in comparison.

However, earlier reconstruction efforts by Lieberman, Laitman and Crelin (discussed in Sect. "Why the long face? The early hominin vocal tract") suffered from methodological issues unknown at the time. Both efforts were based on the assumption that flexion of the skull base provided a reliable basis for inferring the shape of vocal tracts. Subsequent work (D. Lieberman & R. McCarthy, 1999; Fitch & Giedd, 1999) showed that, in modern humans, the tongue, hyoid, and larynx continue to descend after the point of stabilization of basicranial flexion, placing the reliability of reconstruction efforts in jeopardy (P. Lieberman, 2007). While the methodological bases of these older reconstructions have proved unreliable, later work has supported the conclusions drawn therefrom.

More recent estimates of Neanderthal vocal tracts by McCarthy (reported in P. Lieberman & R. McCarthy, 2007, 2015; P. Lieberman, 2007, 2012) suggest that the combination of long faces and short necks could not have accommodated modern human-like vocal tract proportions necessary for the greatest range of signal variability (Carré et al., 1995, 2017). Fitting the dimensions of the human vocal tract (with a "roughly 1:1" relationship between horizontal,  $SVT_H$ , and vertical sections,  $SVT_V$ ) to the *la chappelle*, *la ferrassie*, and *le moustier* skulls effectively places the larynx in the thorax, an anatomical configuration that does not exist in any mammal (P. Lieberman, 2012). In comparison, for all but one (*Předmostí 3*) of eight Late Pleistocene *H. sapiens* specimen examined – including San Teodoro, Cro Magnon, Dolní Věstonice III, Fish Hoek, Grotte des Enfants 6, Hotu 2, and Zhoukoudian UC 101 skulls – SVT reconstructions were found to have been confined to the neck (and not extend to the thorax). Setting aside for the moment that the exact implications of modern human-like SVT proportions may be ambiguous with regard to fluid speech, these results suggest there were likely differences in vocal tract anatomy between Neanderthals and archaeologically modern *H. sapiens*.

Further caution is warranted in interpreting this estimate. The most generous estimates from the McCarthy vocal tract reconstructions give Neanderthals a ~1.3:1  $SVT_H/SVT_V$  proportion. If these estimates were to fall within the extreme ranges observed in modern humans, it would be cause for disputing the relevance of SVT proportions. However, most studies investigating this relationship in modern human are small-sample works, concerned with the ontogenic growth of the vocal tract (D. Lieberman & R. McCarthy, 1999; D. Lieberman et al., 2001; Vorperian et al., 2005, 2009). Moran et al. (2024) provide measurements of  $SVT_H/SVT_V$  proportions for 55 adult speakers (27 females, 28 males). Results showed proportions of 1:1 ( $SD=0.12$ ) for males, and 1.1:1 ( $SD=0.12$ ) for females (reflecting the pubertal secondary descent of the larynx in males). The most disproportionate proportions observed in the sample was an outlier female at 1.24:1. While these data fall short of the McCarthy estimates of Neanderthal SVT proportions, the relative proximity of values suggests caution is warranted. To our knowledge, there is no adult speaker data to support overlap in proportional estimates between modern humans and the Neanderthal specimen studied by R. McCarthy.



That Neanderthal phonemic space was less extensive than that of modern humans was also supported independently by Barney et al. (2012). The only work to have determined that Neanderthal phonetic space was as extensive as that of modern human was performed by Boë et al., (1999, 2002). However, this work has been refuted. The authors' procedure involved the application of an algorithm that preserves the tongue shapes of the adult humans upon which it was based. Applying the same algorithm, any vocal tract would be shown to possess the full range of modern human speech sounds (de Boer & Fitch, 2010), and so results from these simulations are uninformative about the evolution of speech. The same authors have disputed the relevance of the "1:1 proportions" assumption (Badin et al., 2014). These data dispute the model developed by de Boer (2010) by building a model that includes the lips. These estimates remain concentrated on the possibility of production of single phonemes. As such, however, these data do not dispute other speech acoustics experiments that have reached the same conclusions independently of de Boer's efforts.

For example, findings by Carré and colleagues (1995, 2017) were that essentially human proportions were re-invented by an algorithm set to optimize information transmission capacities of a linear 18 cm tube sequence. These experiments have been exhaustively replicated (Carré et al., 2017). Thus, even if Neanderthals were *not* precluded from the full range of speech sounds, the ability to "optimally" navigate the extent of articulatory-acoustic relationships may yet have been reduced. If so, the modern human vocal tract would indeed be uniquely well-adapted to formant-based communication. Regarding Neanderthals, the open question is whether fluid speech (not solely the species' vowel space) is unperturbed when these dimensions are significantly disturbed from the (archaeologically modern human) norm. This more complex question has never been subjected to empirical testing. Moving forward, it will be necessary to explicitly verify any phonetic advantage to fluid coarticulated speech bestowed by a change in proportions per se.

## Putting the Pieces Together: Invention of the Syllable

There is now a range of evidence indicating extant non-human great apes vocalize voluntarily (Lameira & Shumaker, 2019; Lameira et al., 2016) and are capable of learning novel articulatory gestures (Ekström, 2023; Ekström et al., 2024a; Hopkins et al., 2007; Janik & Slater, 2000; Lameira, 2017; Salmi et al., 2022a, 2022b; Wich et al., 2009), with "syllable-like" articulatory cycles observed across primates (Bergman, 2013; Ekström et al., 2024a). While species' anatomy apparently precludes production of the full extent of speech sounds (P. Lieberman et al., 1969, 1972; Stevens & Blumstein, 1975; Stevens, 1989; Takemoto, 2008; Fitch et al., 2016; P. Lieberman, 2017; Ekström, 2024), it evidently allows for a limited set of modern human speech sounds including rudimentary syllables (Ekström et al., 2024a). This prompts the question as to why, given articulatory possibilities seemingly sufficient for a system of syllabic "proto-speech" to evolve, no such communicative system exists in non-human great apes. To elucidate why this may be so, it is necessary to

briefly consider the communicative benefits of human syllabic speech over animal calls.

The cyclical nature of syllabic speech, organized roughly as frame/content (for present purposes equivalent to consonant/vowel) units, lends itself to fortuitous “chunking” of the signal (MacNeilage, 1998), which makes it more readily appreciated and perceived by human listeners (Lieberman & Mattingly, 1985, 1989; Liberman et al., 1967). This is meaningful as human systems of perception and short-term memory are limited in information storage capacity at any one time (Miller, 1956; Shiffrin & Nosofsky, 1994). The syllabic nature of natural-sounding speech counteracts and eases the perceptual and cognitive load of auditory stimulus perception, with signal predictability (Peters et al., 2016; Sussman & Gumenyuk, 2005; Wilsch & Obleser, 2016) and enculturation of speech sounds (Ekström, 2022; Liberman et al., 1967; Lindblom, 1990; McMurray, 2022; Möttönen & Watkins, 2009; Pisoni, 1971) driving reliability of perception (Table 2). The development and maintenance of systems of speech tend toward exploitation of articulatory-acoustic relationships, while subject to economic constraints of production, perception, and contrast (Liljencrants & Lindblom, 1972; Stevens, 1989). To quote Liljencrants and Lindblom (1972, p. 856), “It seems reasonable to suppose that a system [of speech sounds] which has been optimized with respect to communicative efficiency consists of [sounds] that are not only ‘easy to hear’ but also ‘easy to say’.” We might reasonably expect an early system of syllabic speech to abide by the same universal principles.

The first species to have realized syllabic speech would presumably have seized on the opportunity to exploit the rapid phonetic transitions made available by labial consonant-schwa or schwa-labial consonant transitions. The argument for schwa as early syllabic “content” (MacNeilage, 1998) is two-fold. First, while a variety of primates produce vowel-like calls that overlap with distributions of modern human vowels (Boë et al., 2017; Ekström et al., 2023; Grawunder et al., 2022), these appear associated with significant articulatory effort, and as such are likely not conducive to fluid coarticulated speech (Berthommier, 2020; Ekström, 2024). Second, recent data show that formant patterns in chimpanzees producing the human word “mama” are mostly consistent with schwa (Ekström et al., 2024a). To date, this represents the only evidence of truly syllabic utterances by great apes. Unless speech was invented *de novo* in *H. sapiens* (a suggestion we find evolutionarily implausible), vowel sounds available to the first-ever speakers of syllabic spoken language likely did not include the articulatory degrees of freedom afforded by the modern human vocal tract, as observed across human languages today (Wood, 1979, 1986; Stevens, 1989; Lindblom et al., 1983; P. Lieberman, 1984, 2012, 2017; Carré et al., 1995, 2017; Blasi et al., 2019; Moran & McCloy, 2019; Moran et al., 2021; Sato et al., 2023; Ekström, 2024).

The ability to couple phonation to voluntary mandible movements exists in extant great apes (Ekström et al., 2024a; Lameira & Hardus, 2023), and thus likely also in early-Pliocene hominins. The syllabic utterances theoretically available to these early hominins would certainly have included labial consonant–vowel utterances such as [mama], [wawa], and [baba], and combinations such as [mawa] (Ekström et al., 2024a). However, as noted previously, australopiths may have possessed air



**Table 2** Speculative relationship between evolved vocal tract morphology and their possible phonetic consequences

Epoch	Morphology	Behavior
<i>Extant great apes</i>		
	Robust mandible	Possibly adaptation for masticating hard foods, and used in interindividual conflicts
	Laryngeal air sacs likely a derived feature, though subject to extensive variation between species	Function likely bears relationship to social organization, as evidenced by marked sexual dimorphism of the organs in sexually competitive species, but not in multi-male-multi-female living species. Acoustic effects suggested by computational models but disputed by call acoustics data
	Fleshy lips subject to voluntary control	Lip protrusion and rounding employed analogously to that observed in speech
	Flat tongue with bunching capacities largely absent from the tongue body	Tongue retraction inferred for selections of calls, velar sounds and coarticulatory capacity likely reduced
~ 3.9–2.9 mya	Australopiths may have possessed laryngeal air sacs, derived hyoid morphology	Air sacs likely induce “breathiness” and possibly limit range of vowel-like qualities via the introduction of an additional low-frequency formant
	“Ape-like” basicranium	Likely bears a relationship to vocal tract anatomy
	Relatively robust masticatory apparatus	Robustness of the mandible impaired maximum syllable-per-second rate
~ 2.9–1.7 mya	Reduced mandible myosin-heavy fibers	Affected time spent engaging the articulatory complex through masticating
~ 1.9 mya	Gracilization of the mandible, and retraction of the face	Retraction of the face represents a pre-adaptation for speech production anatomy. Improves maximum syllable rate
	Basicranium of <i>H. erectus</i> a midpoint between Australopith and modern human	Possible relationship to pharyngeal dimensions
	Air sacs likely lost in or prior to the common ancestor of <i>H. neanderthalensis</i> and <i>H. sapiens</i>	Loss argued to have improved speech intelligibility, though exact relevance is disputed
~ 500–300 kya	Convincing evidence of the emergence of cumulative culture	May have been meaningful in developing and maintaining culturally pertinent aspects of phonological variation
	Mutation of <i>FOXP2</i>	Exact contribution to language uncertain but lack of gene disrupts normal language use, including oral-motor planning

Table 2 (continued)

Epoch	Morphology	Behavior
~300–50 kya	Human ancestors may have achieved modern vocal tract dimensions, with uniquely orthognathic faces and expansive pharyngeal cavities	The full extent of speech capacities is achieved, possibly reflecting a novel selection pressure for improved speech communication; may also have facilitated more ready exploitation of articulatory-acoustic relationships
~12 kya	Bite size configuration through the introduction of agriculture	Facilitated incorporation of labiodental phonemes into existing phonologie

sacs, in which case their speech would have exhibited a “breathy” quality and the resulting vowel space may have been limited (de Boer, 2012). Tentative observations by Ekström et al. (2023) suggested that orangutan phonation is characterized by seemingly incomplete glottal closure, indicating a lack of signal stability (see also Nishimura et al., 2022; Sato et al., 2023). By the emergence of genus *Homo*, air sacs were – judging by the comparative shape of the hyoid – completely or partially lost (P. Lieberman and R. McCarthy, 2015), freeing the signal from its possible initial breathy constraints. However, syllables such as [mə] (“muh”) offer comparatively rich linguistic possibilities, resulting from their articulatory straightforwardness (as evident from their early appearance in human phonological development), and acoustic and perceptual distinctiveness. At the moment of mouth opening, acoustic energy is redirected from the nasal cavities to the oral cavity, resulting in abrupt change in sound quality. Other labial sounds including [m], [b], and [p] are likewise acoustically distinct, and can be – and some indeed are – straightforwardly articulated with unconfigured vocal tracts such as those exhibited by non-human primates (P. Ekström, 2023; Ekström et al., 2024a; Lameira & Moran, 2023; Lieberman et al., 1992) and human infants (D. Green & Nip, 2010; McCarthy, 1946). An early system of syllabic speech may well also have exploited the apparent acoustic and articulatory distinctions between partial and complete closure of the mouth and lips (e.g., “muh”, cf. “wuh”). As to the issue of why no such system evolved in non-human great apes, we see two main possibilities.

The first possibility is that anatomical-economic costs of speaking, in combination with usefulness of then-existent repertoires of calls, may have been too great to overcome. Anatomical changes such as expansion of the epiglottic space (Sato et al., 2023) and downstream increased flexibility of the pharyngeal tract may thus have been necessary for a system of speech biomechanics to evolve (Ekström, 2024; Jürgens, 2002; Lameira, 2017; MacNeilage, 1998; Takemoto, 2008). Second, more speculatively, time spent masticating by definition precluded time spent otherwise engaged. “Freeing up” of time actively engaging the mandible and vocal tract may thus have allowed the rudimentary vocal control capabilities of non-human great apes (Ekström et al., 2024a; Lameira & Shumaker, 2019) to elaborate and experiment with articulated sounds (Dezecache et al., 2021; Ekström, 2022; Fee & Goldberg, 2011; Lameira et al., 2022; LeBlois et al., 2010; Metfessel, 1935). So established, such an ability may have allowed cultural evolution to enact novel pressures on biological and genetic evolution (Heinrich, 2016; Markov & Markov, 2020) toward clarity, consistency, and signal robustness.

We endorse a combination of models. Anatomical evolution was necessary to produce the full range of human speech sounds—possibly including its most acoustically and perceptually distinct sounds—but also to efficiently navigate the full extent of articulatory-acoustic relationships (Carré et al., 1995, 2017). We have suggested that the anatomy co-opted for speech in human evolution co-evolved with adaptations to dietary change. A dietary shift toward processed (and ultimately cooked) food coincides in the paleoanthropological record with a general reduction in craniofacial size in late australopiths and early *Homo*, illustrating that human ancestors became increasingly “pre-adapted” for speech. The “outsourcing” of food ingestion and digestion (e.g., initially to hand, tool and/or fermentation etc., and

later to fire) by ancestral hominins may have facilitated the reduction and subsequent maintenance of various would-be articulatory morphological elements, including the mandible and midface, facilitating more extensive, elaborative vocal-articulatory communication. With less time spent masticating food, early speakers may have been free to experiment with basic tenets of speech (Planer and Sterelny, 2021). *H. erectus* used stone tools for mechanical processing of food. Behavioral outsourcing of ingestion and digestion may have in turn facilitated both the gracilization of the mandible and subsequent growth of the brain in the *Homo* lineage. Truly cumulative culture may have emerged by the mid-Pleistocene some 500–300 kya, in archaic *Homo*, whereas a fully modern human vocal tract may have emerged as late as 100 kya (P. Lieberman & R. McCarthy, 2007, 2015; P. Lieberman, 2012).

Our tentative view integrates existing and new perspectives on anatomical, neural, social, and cultural evolution. Once established, a rudimentary system of speech sounds may subsequently have enacted novel selection pressures, preserving a novel mutation of the FOXP2 gene variant in Neanderthals and *H. sapiens* (Krause et al., 2007; Kuhlwilms, 2018; Zeberg et al., 2024) and a unique selection pressure for articulate speech in Upper Pleistocene early *H. sapiens*. Following Negus (1949) and Lieberman and colleagues (1972), this last push toward modern articulatory morphology may thus have been driven by pressure for clearer, more robust speech. It will be necessary that future work moves away from static vowel production (P. Berthommier, 2020; Boë et al., 2017; Fitch et al., 2016; Lieberman et al., 1969, 1972) and toward a model capable of explicating the relevance of comparative articulatory morphology for the rapid and fluid production of syllabic speech (Carré et al., 2017; Lindblom, 1983; Lindblom & MacNeilage, 2011; MacNeilage, 1998; Studdert-Kennedy, 1998). Any account of the evolution of speech or phonetic capacities must ultimately reconcile with the enormous suite of changes that characterized the evolution of our unique species. A synthesis of paleoanthropological, archaeological, and phonetic evidence indicates that co-evolution of body-external food ingestion and digestion techniques (including stone toolmaking and butchery), emergence of culture, and increased social consciousness all preceded (and coincided with) the evolution of a modern human vocal tract.

**Acknowledgements** The results of this work and the tools used will be made more widely accessible through the national infrastructure Språkbanken Tal under funding from the Swedish Research Council (2017–00626). AGE, DF and SM received funding through the SNSF (Grant No. PCEFP1\_186841). David S. Strait was further supported by the German Research Foundation (DFG FOR 2237: Project “Words, Bones, Genes, Tools: Tracking Linguistic, Cultural, and Biological Trajectories of the Human Past”).

We thank Richard Wrangham and Björn Lindblom for discussions on the topics presented here, and Marlen Fröhlich for prompting additional clarifications.

**Author Contributions** AGE: Conceptualization, Writing – Original draft, Writing – Editing and review; PG: Writing – Editing and review; WDS: Writing – Editing and review; DF: Writing – Editing and review; RCM: Writing – Editing and review; MT: Writing – Editing and review; CT: Writing – Editing and review; DSS: Writing – Editing and review; JE: Funding acquisition; Writing – Editing and review; SM: Project supervision; Funding acquisition; Writing – Editing and review.

**Funding** Open access funding provided by Royal Institute of Technology. Swedish Research Council, 2017–00626 (Jens Edlund PI). Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, PCEFP1\_186841 (Steven Moran PI).

**Data Availability** Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

## Declarations

**Conflicts of Interest** The authors confirm that we have no conflicts of interest regarding this work.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abdollahzadeh, A., McPherron, S. P., Sandgathe, D. M., Schurr, T. G., Olszewski, D. I., & Dibble, H. L. (2022). Investigating variability in the frequency of fire use in the archaeological record of Late Pleistocene Europe. *Archaeological and Anthropological Sciences*, 14(4), 62. <https://doi.org/10.1007/s12520-022-01526-1>
- Ackermann, H. (2008). Cerebellar contributions to speech production and speech perception: Psycholinguistic and neurobiological perspectives. *Trends in Neurosciences*, 31(6), 265–272. <https://doi.org/10.1016/j.tins.2008.02.011>
- Aiello, L. C., & Wheeler, P. (1995). The expensive-tissue hypothesis: The brain and the digestive system in human and primate evolution. *Current Anthropology*, 36(2), 199–221. <https://doi.org/10.1086/204350>
- Alemseged, Z., Spoor, F., Kimbel, W. H., Bobe, R., Geraads, D., Reed, D., & Wynn, J. G. (2006). A juvenile early hominin skeleton from Dikika, Ethiopia. *Nature*, 443(7109), 296–301. <https://doi.org/10.1038/nature05047>
- Alexander, M. P., Naeser, M. A., & Palumbo, C. L. (1987). Correlations of subcortical lesion sites and aphasia profiles. *Brain*, 110(4), 961–988. <https://doi.org/10.1093/brain/110.4.961>
- Alm, P. A. (2021). The dopamine system and automatization of movement sequences: a review with relevance for speech and stuttering. *Frontiers in Human Neuroscience*, 663. <https://doi.org/10.3389/fnhum.2021.661880>
- Amiez, C., Verstraete, C., Sallet, J., Hadj-Bouziane, F., Ben Hamed, S., Meguerditchian, A., ... & Hopkins, W. D. (2023). The relevance of the unique anatomy of the human prefrontal operculum to the emergence of speech. *Communications Biology*, 6(1), 693. <https://doi.org/10.1038/s42003-023-05066-9>
- Badin, P., Boë, L. J., Sawallis, T. R., & Schwartz, J. L. (2014). Keep the lips to free the larynx: Comments on de Boer's articulatory model (2010). *Journal of Phonetics*, 46, 161–167. <https://doi.org/10.1016/j.wocn.2014.07.002>
- Barkai, R., Rosell, J., Blasco, R., & Gopher, A. (2017). Fire for a reason: Barbecue at middle Pleistocene Qesem cave, Israel. *Current Anthropology*, 58(S16), S314–S328. <https://doi.org/10.1086/691211>
- Barlow, S. M., & Estep, M. (2006). Central pattern generation and the motor infrastructure for suck, respiration, and speech. *Journal of Communication Disorders*, 39(5), 366–380. <https://doi.org/10.1016/j.jcomdis.2006.06.011>
- Barney, A., Martelli, S., Serrurier, A., & Steele, J. (2012). Articulatory capacity of Neanderthals, a very recent and human-like fossil hominin. *Philosophical Transactions of the Royal Society b: Biological Sciences*, 367(1585), 88–102. <https://doi.org/10.1098/rstb.2011.0259>
- Belyk, M., & Brown, S. (2017). The origins of the vocal brain in humans. *Neuroscience & Biobehavioral Reviews*, 77, 177–193. <https://doi.org/10.1016/j.neubiorev.2017.03.014>

- Ben-Dor, M., Gopher, A., Hershkovitz, I., & Barkai, R. (2011). Man the fat hunter: the demise of *Homo erectus* and the emergence of a new hominin lineage in the Middle Pleistocene (ca. 400 kyr) Levant. *PLoS One*, 6(12), e28689. <https://doi.org/10.1371/journal.pone.0028689>
- Benítez-Burraco, A., & Kempe, V. (2018). The emergence of modern languages: Has human self-domestication optimized language transmission? *Frontiers in Psychology*, 9, 551. <https://doi.org/10.3389/fpsyg.2018.00551>
- Bergman, T. J. (2013). Speech-like vocalized lip-smacking in geladas. *Current Biology*, 23(7), R268–R269. <https://doi.org/10.1016/j.cub.2013.02.038>
- Bermejo-Fenoll, A., Panchón-Ruiz, A., & Sánchez del Campo, F. (2019). *Homo sapiens*, Chimpanzees and the Enigma of Language. *Frontiers in Neuroscience*, 13, 558. <https://doi.org/10.3389/fnins.2019.00558>
- Berna, F., Goldberg, P., Horwitz, L. K., Brink, J., Holt, S., Bamford, M., & Chazan, M. (2012). Microstratigraphic evidence of in situ fire in the Acheulean strata of Wonderwerk Cave, Northern Cape province, South Africa. *Proceedings of the National Academy of Sciences*, 109(20), E1215–E1220. <https://doi.org/10.1073/pnas.1117620109>
- Berthommier, F., Boë, L. J., Meguerditchian, A., Sawallis, T., & Captier, G. (2017). Comparative anatomy of the baboon and human vocal tracts: Renewal of methods, data, and hypotheses. In L. J. Boë, J. Fagot, P. Perrier, & J. L. Schwartz (Eds.), *Speech production and perception* (Vol. 4, pp. 101–136). Peter Lang.
- Berthommier, F. (2020). Monkey vocal tracts are not so “speech ready”. In N. H. Bernadoni & L. Bailly, (Eds.), *Proceedings of 12th International Conference on Voice Physiology and Biomechanics* (p. 28). Université Grenoble Alpes.
- Bianchi, S., Reyes, L. D., Hopkins, W. D., Taglialatela, J. P., & Sherwood, C. C. (2016). Neocortical grey matter distribution underlying voluntary, flexible vocalizations in chimpanzees. *Scientific Reports*, 6(1), 34733. <https://doi.org/10.1038/srep34733>
- Blasi, D. E., Moran, S., Moisik, S. R., Widmer, P., Dediu, D., & Bickel, B. (2019). Human sound systems are shaped by post-Neolithic changes in bite configuration. *Science*, 363(6432), eaav3218. <https://doi.org/10.1126/science.aav3218>
- Blumenschine, R. J., & Pobiner, B. L. (2007). Zooarchaeology and the ecology of Oldowan hominin carnivory. In P. S. Ungar (Ed.), *Evolution of the human diet: the known, the unknown, and the unknowable*, 167–190.
- Boë, L. J., Maeda, S., & Heim, J. L. (1999). Neandertal man was not morphologically handicapped for speech. *Evolution of Communication*, 3(1), 49–77. <https://doi.org/10.1075/eoc.3.1.05boe>
- Boë, L. J., Heim, J. L., Honda, K., & Maeda, S. (2002). The potential Neandertal vowel space was as large as that of modern humans. *Journal of Phonetics*, 30(3), 465–484. <https://doi.org/10.1006/jpho.2002.0170>
- Boë, L. J., Berthommier, F., Legou, T., Captier, G., Kemp, C., Sawallis, T. R., ... & Fagot, J. (2017). Evidence of a vocalic proto-system in the baboon (*Papio papio*) suggests pre-hominin speech precursors. *PLoS One*, 12(1), e0169321. <https://doi.org/10.1371/journal.pone.0169321>
- Boëda, E., Geneste, J. M., Griggo, C., Mercier, N., Muhesen, S., Reyss, J. L., ... & Valladas, H. (1999). A Levallois point embedded in the vertebra of a wild ass (*Equus africanus*): hafting, projectiles and Mousterian hunting weapons. *Antiquity*, 73(280), 394–402. <https://doi.org/10.1017/S0003598X0008335>
- Boesch, C., & Boesch, H. (1983). Optimisation of nut-cracking with natural hammers by wild chimpanzees. *Behaviour*, 83(3–4), 265–286. <https://doi.org/10.1163/156853983X00192>
- Boesch, C., & Boesch, H. (1990). Tool use and tool making in wild chimpanzees. *Folia Primatologica*, 54(1–2), 86–99. <https://doi.org/10.1159/000156428>
- Brace, C. L. (1999). An anthropological perspective on “race” and intelligence: The non-clinal nature of human cognitive capabilities. *Journal of Anthropological Research*, 55(2), 245–264. <https://doi.org/10.1086/jar.55.2.3631210>
- Bräuer, J., & Call, J. (2015). Apes produce tools for future use. *American Journal of Primatology*, 77(3), 254–263. <https://doi.org/10.1002/ajp.22341>
- Broadhurst, C. L., Wang, Y., Crawford, M. A., Cunnean, S. C., Parkington, J. E., & Schmidt, W. F. (2002). Brain-specific lipids from marine, lacustrine, or terrestrial food resources: Potential impact on early African *Homo sapiens*. *Comparative Biochemistry and Physiology Part b: Biochemistry and Molecular Biology*, 131(4), 653–673. [https://doi.org/10.1016/S1096-4959\(02\)00002-7](https://doi.org/10.1016/S1096-4959(02)00002-7)

- Brown, S., Yuan, Y., & Belyk, M. (2021). Evolution of the speech-ready brain: The voice/jaw connection in the human motor cortex. *Journal of Comparative Neurology*, 529(5), 1018–1028. <https://doi.org/10.1002/cne.24997>
- Bunn, H. T. (2007). Meat made us human. In P. S. Ungar (Ed.), *Evolution of the human diet: The known, the unknown, and the unknowable* (pp. 191–211). Oxford University Press.
- Burini, R. C., & Leonard, W. R. (2018). The evolutionary roles of nutrition selection and dietary quality in the human brain size and encephalization. *Nutrire*, 43(1), 1–9. <https://doi.org/10.1186/s41110-018-0078-x>
- Bramble, D. M., & Lieberman, D. E. (2004). Endurance running and the evolution of *Homo*. *Nature*, 432(7015), 345–352. <https://doi.org/10.1038/nature03052>
- Cantalupo, C., & Hopkins, W. D. (2001). Asymmetric Broca's area in great apes. *Nature*, 414(6863), 505–505. <https://doi.org/10.1038/35107134>
- Carlson, K. J., Stout, D., Jashashvili, T., De Ruiter, D. J., Tafforeau, P., Carlson, K., & Berger, L. R. (2011). The endocast of MH1, *Australopithecus sediba*. *Science*, 333(6048), 1402–1407. <https://doi.org/10.1126/science.1203922>
- Carmody, R. N., & Wrangham, R. W. (2009). The energetic significance of cooking. *Journal of Human Evolution*, 57(4), 379–391. <https://doi.org/10.1016/j.jhevol.2009.02.011>
- Carré, R., Lindblom, B., & MacNeilage, P. (1995). Acoustic factors in the evolution of the human vocal tract. *Comptes Rendus De L'academie des Sciences Serie II*, 320(9), 471–476.
- Carré, R., Divenyi, P., & Mrayati, M. (2017). Speech: A dynamic process. *De Gruyter*. <https://doi.org/10.1515/9781501502019>
- Cáceres, I., Chelli Cheheb, R., Van der Made, J., Harichane, Z., Boulaghraief, K., & Sahnouni, M. (2023). Assessing the subsistence strategies of the earliest North African inhabitants: Evidence from the Early Pleistocene site of Ain Boucherit (Algeria). *Archaeological and Anthropological Sciences*, 15(6), 87. <https://doi.org/10.1007/s12520-023-01783-8>
- Chivers, D. J., & Hladik, C. M. (1984). Diet and gut morphology in primates. In D. J. Chivers, B. A. Wood, & A. Bilsborough (Eds.), *Food acquisition and processing in primates* (pp. 213–230). Springer.
- Chitra, U., Singh, U., & Venkateswara Rao, P. (1996). Phytic acid, in vitro protein digestibility, dietary fiber, and minerals of pulses as influenced by processing methods. *Plant Foods for Human Nutrition*, 49(4), 307–316. <https://doi.org/10.1007/BF01091980>
- Cieri, R. L., Churchill, S. E., Franciscus, R. G., Tan, J., & Hare, B. (2014). Craniofacial feminization, social tolerance, and the origins of behavioral modernity. *Current Anthropology*, 55(4), 419–443. <https://doi.org/10.1086/677209>
- Clark, G., & Henneberg, M. (2017). *Ardipithecus ramidus* and the evolution of language and singing: An early origin for hominin vocal capability. *Homo*, 68(2), 101–121. <https://doi.org/10.1016/j.jchb.2017.03.001>
- Clark, G., & Henneberg, M. (2021). Cognitive and behavioral modernity in *Homo erectus*: Skull globularity and hominin brain evolution. *Anthropological Review*, 84(4), 467–485. <https://doi.org/10.2478/anre-2021-0030>
- Clark, G., & Henneberg, M. (2022). Interpopulational variation in human brain size: Implications for hominin cognitive phylogeny. *Anthropological Review*, 84(4), 405–429. <https://doi.org/10.5167/uzh-214484>
- Connor, R. C. (2007). Dolphin social intelligence: Complex alliance relationships in bottlenose dolphins and a consideration of selective environments for extreme brain size evolution in mammals. *Philosophical Transactions of the Royal Society b: Biological Sciences*, 362(1480), 587–602. <https://doi.org/10.1098/rstb.2006.1997>
- Corballis, M. C. (2002). *From hand to mouth: The origins of language*. Princeton University Press.
- Cornélio, A. M., de Bittencourt-Navarrete, R. E., de Bittencourt Brum, R., Queiroz, C. M., & Costa, M. R. (2016). Human brain expansion during evolution is independent of fire control and cooking. *Frontiers in Neuroscience*, 10, 167. <https://doi.org/10.3389/fnins.2016.00167>
- Crelin, E. S. (1987). *The human vocal tract: Anatomy, function, development, and evolution*. Vantage Press.
- Crelin, E. S. (1989). The skulls of our ancestors: Implications regarding speech, language, and conceptual thought evolution. *Journal of Voice*, 3(1), 18–23. [https://doi.org/10.1016/S0892-1997\(89\)80117-1](https://doi.org/10.1016/S0892-1997(89)80117-1)
- Crockford, C., Herlinger, I., Vigilant, L., & Boesch, C. (2004). Wild chimpanzees produce group-specific calls: A case for vocal learning? *Ethology*, 110(3), 221–243. <https://doi.org/10.1111/j.1439-0310.2004.00968.x>



- de Boer, B. (2010). Investigating the acoustic effect of the descended larynx with articulatory models. *Journal of Phonetics*, 38(4), 679–686. <https://doi.org/10.1016/j.wocn.2010.10.003>
- de Boer, B. (2012). Loss of air sacs improved hominin speech abilities. *Journal of Human Evolution*, 62(1), 1–6. <https://doi.org/10.1016/j.jhevol.2011.07.007>
- de Boer, B., & Fitch, T. W. (2010). Computer models of vocal tract evolution: An overview and critique. *Adaptive Behavior*, 18(1), 36–47. <https://doi.org/10.1177/1059712309350972>
- de Boer, B., Wich, S. A., Hardus, M. E., & Lameira, A. R. (2015). Acoustic models of orangutan hand-assisted alarm calls. *The Journal of Experimental Biology*, 218(6), 907–914. <https://doi.org/10.1242/jeb.110577>
- Delattre, P. C., Liberman, A. M., & Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants. *The Journal of the Acoustical Society of America*, 27(4), 769–773. <https://doi.org/10.1121/1.1908024>
- Demes, B., & Creel, N. (1988). Bite force, diet, and cranial morphology of fossil hominids. *Journal of Human Evolution*, 17(7), 657–670. [https://doi.org/10.1016/0047-2484\(88\)90023-1](https://doi.org/10.1016/0047-2484(88)90023-1)
- DeSilva, J. M., Traniello, J. F., Claxton, A. G., & Fannin, L. D. (2021). When and why did human brains decrease in size? A new change-point analysis and insights from brain evolution in ants. *Frontiers in Ecology and Evolution*, 712. <https://doi.org/10.3389/fevo.2021.742639>
- Dew, J. L. (2005). Foraging, food choice, and food processing by sympatric ripe-fruit specialists: *Lagothrix lagotricha poeppigii* and *Ateles belzebuth belzebuth*. *International Journal of Primatology*, 26(5), 1107–1135. <https://doi.org/10.1007/s10764-005-6461-5>
- Dezecache, G., Zuberbühler, K., Davila-Ross, M., & Dahl, C. D. (2021). Flexibility in wild infant chimpanzee vocal behavior. *Journal of Language Evolution*, 6(1), 37–53. <https://doi.org/10.1093/jole/lzaa009>
- Diehl, R. L. (1989). Remarks on Stevens' quantal theory of speech. *Journal of Phonetics*, 17(1–2), 71–78. [https://doi.org/10.1016/S0095-4470\(19\)31524-4](https://doi.org/10.1016/S0095-4470(19)31524-4)
- Diehl, R. L. (2008). Acoustic and auditory phonetics: The adaptive design of speech sound systems. *Philosophical Transactions of the Royal Society b: Biological Sciences*, 363(1493), 965–978. <https://doi.org/10.1098/rstb.2007.2153>
- Diogo, R., Richmond, B. G., & Wood, B. (2012). Evolution and homologies of primate and modern human hand and forearm muscles, with notes on thumb movements and tool use. *Journal of Human Evolution*, 63(1), 64–78. <https://doi.org/10.1016/j.jhevol.2012.04.001>
- Dorman, M. F., Studdert-Kennedy, M., & Raphael, L. J. (1977). Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception & Psychophysics*, 22, 109–122. <https://doi.org/10.3758/BF03198744>
- Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience*, 22(1), 567–631. <https://doi.org/10.1146/annurev.neuro.22.1.567>
- Dronkers, N. F., Plaisant, O., Iba-Zizen, M. T., & Cabanis, E. A. (2007). Paul Broca's historic cases: High resolution MR imaging of the brains of Leborgne and Lelong. *Brain*, 130(5), 1432–1441. <https://doi.org/10.1093/brain/awm042>
- Du, A., Zipkin, A. M., Hatala, K. G., Renner, E., Baker, J. L., Bianchi, S., ... & Wood, B. A. (2018). Pattern and process in hominin brain size evolution are scale-dependent. *Proceedings of the Royal Society B: Biological Sciences*, 285(1873), 20172738. <https://doi.org/10.1098/rspb.2017.2738>
- Dunbar, R. I., & Shultz, S. (2007). Evolution in the social brain. *Science*, 317(5843), 1344–1347. <https://doi.org/10.1126/science.1145463>
- Dunbar, R. I. (2014). How conversations around campfires came to be. *Proceedings of the National Academy of Sciences*, 111(39), 14013–14014. <https://doi.org/10.1073/pnas.1416382111>
- Ekström, A. G. (2022). Motor constellation theory: A model of infants' phonological development. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.996894>
- Ekström, A. G. (2023). Viki's First Words: A Comparative Phonetics Case Study. *International Journal of Primatology*, 44, 249–253. <https://doi.org/10.1007/s10764-023-00350-1>
- Ekström, A. G., & Edlund, J. (2023a). Evolution of the tongue and emergence of speech biomechanics. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2023.1150778>
- Ekström, A. G., & Edlund, J. (2023b). Sketches of chimpanzee (*Pan troglodytes*) hoo's: Vowels by any other name? *Primates*. <https://doi.org/10.1007/s10329-023-01107-3>
- Ekström, A. G., Moran, S., Sundberg, J., & Lameira, A. R. (2023). PREQUEL: Supervised phonetic approaches to analyses of great ape quasi-vowels. In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS 2023)*, 3076–3080, Prague, Czech Republic. <https://doi.org/10.31234/osf.io/8aeh4>



- Ekström, A. G. (2024). Correcting the record: Phonetic potential of primate vocal tracts and the legacy of Philip Lieberman (1934–2022). *American Journal of Primatology*, 86(8), e23637. <https://doi.org/10.1002/ajp.23637>
- Ekström, A. G., Gannon, C., Edlund, J., Moran, S., & Lameira, A. R. (2024a). Chimpanzee utterances refute purported missing links for novel vocalizations and syllabic speech. *Scientific Reports*, 14, 17135. <https://doi.org/10.1038/s41598-024-67005-w>
- Ekström, A. G., Cros Vila, L., Schötz, S., & Edlund, J. (2024b). A single formant explicates the ubiquity of “meow”. In M. Miron, & R. Marxer (Eds.), *Proceedings of the 4th international workshop on vocal interactivity in-and-between humans, animals and robots: VIHAR 2024*, 74–79.
- Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S., Wiebe, V., Kitano, T., ... & Pääbo, S. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. *Nature*, 418(6900), 869–872. <https://doi.org/10.1038/nature01025>
- Everett, C. (2017). Yawning at the dawn of speech: A closer look at monkey formant space. Retrieved from: [http://www.calebeverett.org/uploads/4/2/6/5/4265482/commentary\\_on\\_fitch\\_et\\_al..pdf](http://www.calebeverett.org/uploads/4/2/6/5/4265482/commentary_on_fitch_et_al..pdf)
- Falk, D. (1975). Comparative anatomy of the larynx in man and the chimpanzee: Implications for language in Neanderthal. *American Journal of Physical Anthropology*, 43(1), 123–132. <https://doi.org/10.1002/ajpa.1330430116>
- Fant, G. (1960). *The acoustic theory of speech production*. Mouton.
- Fee, M. S., & Goldberg, J. H. (2011). A hypothesis for basal ganglia-dependent reinforcement learning in the songbird. *Neuroscience*, 198, 152–170. <https://doi.org/10.1016/j.neuroscience.2011.09.069>
- Fernández-Jalvo, Y., Tormo, L., Andrews, P., & Marin-Monfort, M. D. (2018). Taphonomy of burnt bones from Wonderwerk Cave (South Africa). *Quaternary International*, 495, 19–29. <https://doi.org/10.1016/j.quaint.2018.05.028>
- Fisher, S. E., & Scharff, C. (2009). FOXP2 as a molecular window into speech and language. *Trends in Genetics*, 25(4), 166–177. <https://doi.org/10.1016/j.tig.2009.03.002>
- Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, 106(3), 1511–1522. <https://doi.org/10.1121/1.427148>
- Fitch, W. T. (2000). The phonetic potential of nonhuman vocal tracts: Comparative cineradiographic observations of vocalizing animals. *Phonetica*, 57(2–4), 205–218. <https://doi.org/10.1159/000028474>
- Fitch, W. T., De Boer, B., Mathur, N., & Ghazanfar, A. A. (2016). Monkey vocal tracts are speech-ready. *Science Advances*, 2(12), e1600723. <https://doi.org/10.1126/sciadv.1600723>
- Friedrichs, D., Maurer, D., Rosen, S., & Dellwo, V. (2017). Vowel recognition at fundamental frequencies up to 1 kHz reveals point vowels as acoustic landmarks. *The Journal of the Acoustical Society of America*, 142(2), 1025–1033. <https://doi.org/10.1121/1.4998706>
- Friedrichs, D. & Dellwo, V. (2022). Are temporal features of voice identity influenced by jaw size? *Proceedings of the 1st Interdisciplinary Voice Identity Conference* (p. 10). University of Zurich.
- Friedrichs, D. & Dellwo, V. (2023). Reorganization of the auditory-perceptual space across the human vocal range. In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 560–560). Guarant International.
- Furness, W. H. (1916). Observations on the mentality of chimpanzees and orang-utans. *Proceedings of the American Philosophical Society*, 55(3), 281–290. <https://www.jstor.org/stable/984118>
- Gannon, P. J., Holloway, R. L., Broadfield, D. C., & Braun, A. R. (1998). Asymmetry of chimpanzee planum temporale: Humanlike pattern of Wernicke’s brain language area homolog. *Science*, 279(5348), 220–222. <https://doi.org/10.1126/science.279.5348.220>
- Gärdenfors, P., Brinck, I., & Osvath, M. (2012). The tripod effect: Co-evolution of cooperation, cognition and communication. In T. Schilhäb, F. Stjernfelt & T. Deacon (Eds.), *The symbolic species evolved*. Springer, 193–222.
- Gärdenfors, P., & Högborg, A. (2017). The archaeology of teaching and the evolution of *Homo docens*. *Current Anthropology*, 58(2), 188–208. <https://doi.org/10.1086/691178>
- Gärdenfors, P., & Lombard, M. (2018). Causal cognition, force dynamics and early hunting technologies. *Frontiers in Psychology*, 9, 87. <https://doi.org/10.3389/fpsyg.2018.00087>
- Gardner, R. A., Gardner, B. T., & Van Cantfort, T. E. (Eds.). (1989). *Teaching sign language to chimpanzees*. Suny Press.
- Gay, T. (1974). A cinefluorographic study of vowel production. *Journal of Phonetics*, 2(4), 255–266. [https://doi.org/10.1016/S0095-4470\(19\)31296-3](https://doi.org/10.1016/S0095-4470(19)31296-3)

- Ghazanfar, A. A., & Takahashi, D. Y. (2014). The evolution of speech: Vision, rhythm, cooperation. *Trends in Cognitive Sciences*, 18(10), 543–553. <https://doi.org/10.1016/j.tics.2014.06.004>
- Goodall, J. (1986). The chimpanzees of Gombe: Patterns of behavior. *Cambridge Mass.*
- Goldberg, P., Rhodes, S. E., & Chazan, M. (2023). Geological and Archeological Insight into Site Formation Processes and Acheulean Occupation at Wonderwerk Cave, Northern Cape Province. *South Africa. Journal of Paleolithic Archaeology*, 6(1), 33. <https://doi.org/10.1007/s41982-023-00157-9>
- Goncharova, M., Jadoul, Y., Reichmuth, C., Fitch, W. T., & Ravignani, A. (2024). Vocal tract dynamics shape the formant structure of conditioned vocalizations in a harbor seal. *Annals of the New York Academy of Sciences*. <https://doi.org/10.1111/nyas.15189>
- Gott, B. (2002). Fire-making in Tasmania: Absence of evidence is not evidence of absence. *Current Anthropology*, 43(4), 650–656. <https://doi.org/10.1086/342430>
- Gowlett, J. A., & Wrangham, R. W. (2013). Earliest fire in Africa: towards the convergence of archaeological evidence and the cooking hypothesis. *Azania: Archaeological Research in Africa*, 48(1), 5–30. <https://doi.org/10.1080/0067270X.2012.756754>
- Gowlett, J. A. (2016). The discovery of fire by humans: A long and convoluted process. *Philosophical Transactions of the Royal Society b: Biological Sciences*, 371(1696), 20150164. <https://doi.org/10.1098/rstb.2015.0164>
- Gracco, V. L., & Abbs, J. H. (1988). Central patterning of speech movements. *Experimental Brain Research*, 71(3), 515–526. <https://doi.org/10.1007/BF00248744>
- Grawunder, S., Uomini, N., Samuni, L., Bortolato, T., Girard-Buttoz, C., Wittig, R. M., & Crockford, C. (2022). Chimpanzee vowel-like sounds and voice quality suggest formant space expansion through the hominoid lineage. *Philosophical Transactions of the Royal Society B*, 377(1841), 20200455. <https://doi.org/10.1098/rstb.2020.0455>
- Green, J. R., & Nip, I. S. (2010). Some organization principles in early speech development. In B. Maaseen and PHHM van Lieshout (eds.), *Speech Motor Control* 10, 171–188. <https://doi.org/10.1093/acprof:oso/9780199235797.003.0010>
- Guenther, F. H. (2016). *Neural control of speech*. MIT Press.
- Gunz, P., Neubauer, S., Falk, D., Tafforeau, P., Le Cabec, A., Smith, T. M., ... & Alemseged, Z. (2020). Australopithecus afarensis endocrasts suggest ape-like brain organization and prolonged brain growth. *Science advances*, 6(14), eaaz4729. <https://doi.org/10.1126/sciadv.aaz4729>
- Harmand, S., Lewis, J. E., Feibel, C. S., Lepre, C. J., Prat, S., Lenoble, A., ... & Roche, H. (2015). 3.3-million-year-old stone tools from Lomekwi 3, West Turkana, Kenya. *Nature*, 521(7552), 310–315. <https://doi.org/10.1038/nature14464>
- Hayden, B. (2014). *The power of feasts: From prehistory to the present*. Cambridge University Press.
- Hayes, K. J., & Hayes, C. (1951). The intellectual development of a home-raised chimpanzee. *Proceedings of the American Philosophical Society*, 95(2), 105–109. <https://www.jstor.org/stable/3143327>
- Heldstab, S. A., Kosonen, Z. K., Koski, S. E., Burkart, J. M., van Schaik, C. P., & Isler, K. (2016). Manipulation complexity in primates coevolved with brain size and terrestriality. *Scientific Reports*, 6(1), 24528. <https://doi.org/10.1038/srep24528>
- Heldstab, S. A., Isler, K., Schuppli, C., & van Schaik, C. P. (2020). When ontogeny recapitulates phylogeny: Fixed neurodevelopmental sequence of manipulative skills among primates. *Science Advances*, 6(30), eabb4685. <https://doi.org/10.1126/sciadv.abb4685>
- Henke, W. L. (1966). Dynamic articulatory model of speech production using computer simulation. [Doctoral Thesis]. Massachusetts Institute of Technology.
- Heimbauer, L. A., Beran, M. J., & Owren, M. J. (2011). A chimpanzee recognizes synthetic speech with significantly reduced acoustic cues to phonetic content. *Current Biology*, 21(14), 1210–1214. <https://doi.org/10.1016/j.cub.2011.06.007>
- Heimbauer, L. A., Beran, M. J., & Owren, M. J. (2021). A chimpanzee recognizes varied acoustical versions of sine-wave and noise-vocoded speech. *Animal Cognition*, 24, 843–854. <https://doi.org/10.1007/s10071-021-01478-4>
- Heinzelin, J. D., Clark, J. D., White, T., Hart, W., Renne, P., WoldeGabriel, G., ... & Vrba, E. (1999). Environment and behavior of 2.5-million-year-old Bouri hominids. *Science*, 284(5414), 625–629. <https://doi.org/10.1126/science.284.5414.625>
- Herndon, J. G., Tigges, J., Anderson, D. C., Klumpp, S. A., & McClure, H. M. (1999). Brain weight throughout the life span of the chimpanzee. *Journal of Comparative Neurology*, 409(4), 567–572. [https://doi.org/10.1002/\(SICI\)1096-9861\(19990712\)409:4%3c567::AID-CNE4%3e3.0.CO;2-J](https://doi.org/10.1002/(SICI)1096-9861(19990712)409:4%3c567::AID-CNE4%3e3.0.CO;2-J)

- Herries, I.R., Adams, J.W., Baker, S., Joannas-Boyau, R., Boschian, G., Mallet T., Murszewski, A., Pickering, R., Caruana, M., Denham, T., Edwards, T.R., Hellstrom, J., Leece, A., Martin, J., Moggi-Cecchi, J., Mokobane, S., Penzo-Kajewski, P., Rovinsky, D., Stammers, R., Strait, D.S., Wilson, C., Woodhead, J., Menter, C. (2020). Contemporaneity of *Australopithecus*, *Paranthropus* and early *Homo erectus* in South Africa. *Science*, 386. <https://doi.org/10.1126/science.aaw7293>
- Hewitt, G., MacLarnon, A., & Jones, K. E. (2002). The functions of laryngeal air sacs in primates: A new hypothesis. *Folia Primatologica*, 73(2–3), 70–94. <https://doi.org/10.1159/000064786>
- Hill, K., Boesch, C., Goodall, J., Pusey, A., Williams, J., & Wrangham, R. (2001). Mortality rates among wild chimpanzees. *Journal of Human Evolution*, 40(5), 437–450. <https://doi.org/10.1006/jhev.2001.0469>
- Hill, H., & Beaudet, A. (2023). Brain evolution and language: A comparative 3D analysis of Wernicke's area in extant and fossil hominids. *Progress in Brain Research*, 275, 117–142. <https://doi.org/10.1016/bs.pbr.2022.12.001>
- Hladik, C. M., Chivers, D. J., & Pasquet, P. (1999). On diet and gut size in non-human primates and humans: Is there a relationship to brain size? *Current Anthropology*, 40(5), 695–697. <https://doi.org/10.1086/300099>
- Hlubik, S., Cutts, R., Braun, D. R., Berna, F., Feibel, C. S., & Harris, J. W. (2019). Hominin fire use in the Okote member at Koobi Fora, Kenya: New evidence for the old debate. *Journal of Human Evolution*, 133, 214–229. <https://doi.org/10.1016/j.jhevol.2019.01.010>
- Hodgson, J. C., & Hudson, J. M. (2018). Speech lateralization and motor control. *Progress in Brain Research*, 238, 145–178. <https://doi.org/10.1016/bs.pbr.2018.06.009>
- Hohmann, G., & Fruth, B. (2007). New records on prey capture and meat eating by bonobos at Lui Kotale, Salonga National Park. *Democratic Republic of Congo. Folia Primatologica*, 79(2), 103–110. <https://doi.org/10.1159/000110679>
- Holloway, R. L. J. (1997). Brain evolution. In R. Dulbecco (Ed.), *Encyclopedia of Human Biology* (pp. 1338–1345). Academic Press.
- Holloway, R. L., Clarke, R. J., & Tobias, P. V. (2004). Posterior lunule sulcus in *Australopithecus africanus*: Was Dart right? *Comptes Rendus Palevol*, 3(4), 287–293. <https://doi.org/10.1016/j.crpv.2003.09.030>
- Hopkins, W. D., Taglialatela, J. P., & Leavens, D. A. (2007). Chimpanzees differentially produce novel vocalizations to capture the attention of a human. *Animal Behaviour*, 73(2), 281–286. <https://doi.org/10.1016/j.anbehav.2006.08.004>
- Humphrey, L. T., Dean, M. C., & Stringer, C. B. (1999). Morphological variation in great ape and modern human mandibles. *The Journal of Anatomy*, 195(4), 491–513. <https://doi.org/10.1046/j.1469-7580.1999.19540491.x>
- Isler, K., Kirk, E. C., Miller, J. M., Albrecht, G. A., Gelvin, B. R., & Martin, R. D. (2008). Endocranial volumes of primate species: Scaling analyses using a comprehensive and reliable data set. *Journal of Human Evolution*, 55(6), 967–978. <https://doi.org/10.1016/j.jhevol.2008.08.004>
- Isler, K., & van Schaik, C. P. (2012). Allomaternal care, life history and brain size evolution in mammals. *Journal of Human Evolution*, 63(1), 52–63. <https://doi.org/10.1016/j.jhevol.2012.03.009>
- Iwasaki, S. I., Yoshimura, K., Shindo, J., & Kageyama, I. (2019). Comparative morphology of the primate tongue. *Annals of Anatomy-Anatomischer Anzeiger*, 223, 19–31. <https://doi.org/10.1016/j.aanat.2019.01.008>
- Izumi, A., & Kojima, S. (2004). Matching vocalizations to vocalizing faces in a chimpanzee (*Pan troglodytes*). *Animal Cognition*, 7, 179–184. <https://doi.org/10.1007/s10071-004-0212-4>
- Jacobs, I., von Bayern, A. M., & Osvath, M. (2021). Tools and food on heat lamps: Pyrocognitive sparks in New Caledonian crows? *Behaviour*, 159(6), 591–602. <https://doi.org/10.1163/1568539X-bja10138>
- Jacob-Friesen, K. H. (1956). Eiszeitliche Elefantenjäger in der Lüneburger Heide. *Jahrbuch des Römisch-Germanischen Zentralmuseums Mainz*, 3, 1–22. <https://doi.org/10.11588/jrgzm.1956.0.32797>
- James, S. R., Dennell, R. W., Gilbert, A. S., Lewis, H. T., Gowlett, J. A. J., Lynch, T. F., ... & James, S. R. (1989). Hominid use of fire in the Lower and Middle Pleistocene: A review of the evidence [and comments and replies]. *Current Anthropology*, 30(1), 1–26. <https://doi.org/10.1086/203705>
- Janik, V. M., & Slater, P. J. (2000). The different roles of social learning in vocal communication. *Animal Behaviour*, 60(1), 1–11. <https://doi.org/10.1006/anbe.2000.1410>
- Johansson, S. (2015). Language abilities in Neanderthals. *Annual Review of Linguistics*, 1(1), 311–332. <https://doi.org/10.1146/annurev-linguist-030514-124945>
- Johnson-Frey, S. H. (2003). What's so special about human tool use? *Neuron*, 39(2), 201–204.

- Jolly, C. J. (1987). The seed-eaters: A new model of hominid differentiation based on a baboon analogy. In R. I. Clochan and J. G. Fleagle (Eds.), *Primate evolution and human origins*. Routledge.
- Jones, M. (2007). *Feast: Why humans share food*. Oxford University Press.
- Jürgens, U. (2002). Neural pathways underlying vocal control. *Neuroscience & Biobehavioral Reviews*, 26(2), 235–258. [https://doi.org/10.1016/S0149-7634\(01\)00068-9](https://doi.org/10.1016/S0149-7634(01)00068-9)
- Jurmain, R., Kilgore, L., & Trevathan, W. (2005). *Introduction to physical anthropology*. Thomson Wadsworth.
- Karakostis, F. A., Hotz, G., Tourloukis, V., & Harvati, K. (2018). Evidence for precision grasping in Neandertal daily activities. *Science Advances*, 4(9), eaat2369. <https://doi.org/10.1126/sciadv.aat2369>
- Karakostis, F. A., Haeufle, D., Anastopoulou, I., Moraitis, K., Hotz, G., Tourloukis, V., & Harvati, K. (2021). Biomechanics of the human thumb and the evolution of dexterity. *Current Biology*, 31(6), 1317–1325. <https://doi.org/10.1016/j.cub.2020.12.041>
- Kataria, A., & Chauhan, B. M. (1988). Contents and digestibility of carbohydrates of mung beans (*Vigna radiata* L.) as affected by domestic processing and cooking. *Plant Foods for Human Nutrition*, 38(1), 51–59. <https://doi.org/10.1007/BF01092310>
- Katz, D. C., Grote, M. N., & Weaver, T. D. (2017). Changes in human skull morphology across the agricultural transition are consistent with softer diets in preindustrial farming groups. *Proceedings of the National Academy of Sciences*, 114(34), 9050–9055. <https://doi.org/10.1073/pnas.1702586114>
- Keeley, L. H. (1980). *Experimental determination of stone tool uses: A microwear analysis*. University of Chicago Press.
- Kewley-Port, D. (1982). Measurement of formant transitions in naturally produced stop consonant–vowel syllables. *The Journal of the Acoustical Society of America*, 72(2), 379–389. <https://doi.org/10.1121/1.388081>
- Koebnick, C., Strassner, C., Hoffmann, I., & Leitzmann, C. (1999). Consequences of a long-term raw food diet on body weight and menstruation: Results of a questionnaire survey. *Annals of Nutrition and Metabolism*, 43(2), 69–79. <https://doi.org/10.1159/000012770>
- Kojima, S., & Kiritani, S. (1989). Vocal-auditory functions in the chimpanzee: Vowel perception. *International Journal of Primatology*, 10, 199–213. <https://doi.org/10.1007/BF02735200>
- Kojima, S., Tatsumi, I. F., Kiritani, S., & Hirose, H. (1989). Vocal-auditory functions of the chimpanzee: Consonant perception. *Human Evolution*, 4, 403–416. <https://doi.org/10.1007/BF02436436>
- Kojima, S. (1990). Comparison of auditory functions in the chimpanzee and human. *Folia Primatologica*, 55(2), 62–72. <https://doi.org/10.1159/000156501>
- Kortüm, H.-H., & Heinze, J. (Eds.). (2013). *Aggression in Humans and Other Primates: Biology, Psychology*. De Gruyter.
- Krause, J., Lalueza-Fox, C., Orlando, L., Enard, W., Green, R. E., Burbano, H. A., ... & Pääbo, S. (2007). The derived FOXP2 variant of modern humans was shared with Neandertals. *Current Biology*, 17(21), 1908–1912. <https://doi.org/10.1016/j.cub.2007.10.008>
- Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *The Journal of the Acoustical Society of America*, 100(4), 2425–2438. <https://doi.org/10.1121/1.417951>
- Kuhlwillm, M. (2018). The evolution of FOXP2 in the light of admixture. *Current Opinion in Behavioral Sciences*, 21, 120–126. <https://doi.org/10.1016/j.cobeha.2018.04.006>
- Lacruz, R. S., Stringer, C. B., Kimbel, W. H., Wood, B., Harvati, K., O'Higgins, P., Bromage, T. B., & Arsuaga, J.-L. (2019). The evolutionary history of the human face. *Nature Ecology & Evolution*, 3, 726–736. <https://doi.org/10.1038/s41559-019-0865-7>
- Lahiff, N. J., Slocombe, K. E., Taglialatela, J., Dellwo, V., & Townsend, S. W. (2022). Degraded and computer-generated speech processing in a bonobo. *Animal Cognition*, 25(6), 1393–1398. <https://doi.org/10.1007/s10071-022-01621-9>
- Laird, M. F., Vogel, E. R., & Pontzer, H. (2016). Chewing efficiency and occlusal functional morphology in modern humans. *Journal of Human Evolution*, 93, 1–11. <https://doi.org/10.1016/j.jhevol.2015.11.005>
- Laitman, J. T., Heimbuch, R. C., & Crelin, E. S. (1979). The basicranium of fossil hominids as an indicator of their upper respiratory systems. *American Journal of Physical Anthropology*, 51(1), 15–33. <https://doi.org/10.1002/ajpa.1330510103>
- Laitman, J. T., & Heimbuch, R. C. (1982). The basicranium of Plio-Pleistocene hominids as an indicator of their upper respiratory systems. *American Journal of Physical Anthropology*, 59(3), 323–343. <https://doi.org/10.1002/ajpa.1330590315>

- Laitman, J. T. (1983). The evolution of the hominid upper respiratory system and implications for the origins of speech. In E. de Grolier (Ed.), *Glossogenetics: The Origin and Evolution of Language* (pp. 63–90). Harwood Academic Publishers.
- Lameira, A. R., Hardus, M. E., Kowalsky, B., de Vries, H., Spruijt, B. M., Sterck, E. H., ... & Wich, S. A. (2013). Orangutan (*Pongo* spp.) whistling and implications for the emergence of an open-ended call repertoire: A replication and extension. *The Journal of the Acoustical Society of America*, 134(3), 2326–2335. <https://doi.org/10.1121/1.4817929>
- Lameira, A. R., Hardus, M. E., Bartlett, A. M., Shumaker, R. W., Wich, S. A., & Menken, S. B. (2015). Speech-like rhythm in a voiced and voiceless orangutan call. *PLoS ONE*, 10(1), e116136. <https://doi.org/10.1371/journal.pone.0116136>
- Lameira, A. R., Hardus, M. E., Mielke, A., Wich, S. A., & Shumaker, R. W. (2016). Vocal fold control beyond the species-specific repertoire in an orang-utan. *Scientific Reports*, 6(1), 30315. <https://doi.org/10.1038/srep30315>
- Lameira, A. R. (2017). Bidding evidence for primate vocal learning and the cultural substrates for speech evolution. *Neuroscience & Biobehavioral Reviews*, 83, 429–439. <https://doi.org/10.1016/j.neubiorev.2017.09.021>
- Lameira, A. R., & Shumaker, R. W. (2019). Orangutans show active voicing through a membrano-phone. *Scientific Reports*, 9(1), 1–6. <https://doi.org/10.1038/s41598-019-48760-7>
- Lameira, A. R., Santamaría-Bonfil, G., Galeone, D., Gamba, M., Hardus, M. E., Knott, C. D., ... & Wich, S. A. (2022). Sociality predicts orangutan vocal phenotype. *Nature Ecology & Evolution*, 6(5), 644–652. <https://doi.org/10.1038/s41559-022-01689-z>
- Lameira, A. R., & Moran, S. (2023). Life of p: A consonant older than speech. *BioEssays*, 45(4), 2200246. <https://doi.org/10.1002/bies.202200246>
- Lameira, A. R., & Hardus, M. E. (2023). Wild orangutans can simultaneously use two independent vocal sound sources similarly to songbirds and human beatboxers. *PNAS Nexus*, 2(6), pgad182. <https://doi.org/10.1093/pnasnexus/pgad182>
- Larsen, C. S. (2003). Animal source foods and human health during evolution. *The Journal of Nutrition*, 133(11), 3893S–3897S. <https://doi.org/10.1093/jn/133.11.3893S>
- Lashley, K. S. (1930). Basic neural mechanisms in behavior. *Psychological Review*, 37(1), 1. <https://doi.org/10.1037/h0074134>
- Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior; the Hixon Symposium* (pp. 112–146). Wiley.
- Leach, H. (2003). Human domestication reconsidered. *Current Anthropology*, 44(3), 349–368. <https://doi.org/10.1086/368119>
- Leakey, L., Tobias, P., & Napier, J. (1964). A new species of the genus *Homo* from Olduvai Gorge. *Nature*, 202, 7–9. <https://doi.org/10.1038/202007a0>
- Leakey, L. S. (1966). *Homo habilis*, *Homo erectus* and the Australopithecines. *Nature*, 209, 1279–1281. <https://doi.org/10.1038/2091279a0>
- Leblois, A., Wendel, B. J., & Perkel, D. J. (2010). Striatal dopamine modulates basal ganglia output and regulates social context-dependent behavioral variability through D1 receptors. *Journal of Neuroscience*, 30(16), 5730–5743. <https://doi.org/10.1523/JNEUROSCI.5974-09.2010>
- Lee, S. H. (2005). Patterns of size sexual dimorphism in *Australopithecus afarensis*: Another look. *Homo*, 56(3), 219–232. <https://doi.org/10.1016/j.jchb.2005.07.001>
- Leigh, S. R. (1992). Cranial capacity evolution in *Homo erectus* and early *Homo sapiens*. *American Journal of Physical Anthropology*, 87(1), 1–13. <https://doi.org/10.1002/ajpa.1330870102>
- Lenneberg, E. H. (1967). The biological foundations of language. *Hospital Practice*, 2(12), 59–67.
- Leroux, M., Bosshard, A. B., Chandia, B., Manser, A., Zuberbühler, K., & Townsend, S. W. (2021). Chimpanzees combine pant hoots with food calls into larger structures. *Animal Behaviour*, 179, 41–50. <https://doi.org/10.1016/j.anbehav.2021.06.026>
- Leroux, M., Schel, A. M., Wilke, C., Chandia, B., Zuberbühler, K., Slocombe, K. E., & Townsend, S. W. (2023). Call combinations and compositional processing in wild chimpanzees. *Nature Communications*, 14(1), 2225. <https://doi.org/10.1038/s41467-023-37816-y>
- Leonard, W. R., Robertson, M. L., & Snodgrass, J. J. (2007). Energetic models of human nutritional evolution. In P. S. Ungar (Ed.), *Evolution of the human diet: The known, the unknown, and the unknowable* (pp. 344–362). Oxford University Press.
- Levelt, W. J. (1999). Models of word production. *Trends in Cognitive Sciences*, 3(6), 223–232. [https://doi.org/https://doi.org/10.1016/S1364-6613\(99\)01319-4](https://doi.org/https://doi.org/10.1016/S1364-6613(99)01319-4)



- Lewis, J. E., & Harmand, S. (2016). An earlier origin for stone tool making: Implications for cognitive evolution and the transition to Homo. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1698), 20150233. <https://doi.org/10.1098/rstb.2015.0233>
- Lieberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74(6), 431–461. <https://doi.org/10.1037/h0020279>
- Lieberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1), 1–36. [https://doi.org/10.1016/0010-0277\(85\)90021-6](https://doi.org/10.1016/0010-0277(85)90021-6)
- Lieberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science*, 243(4890), 489–494. <https://doi.org/10.1126/science.2643163>
- Lieberman, D. E., & McCarthy, R. C. (1999). The ontogeny of cranial base angulation in humans and chimpanzees and its implications for reconstructing pharyngeal dimensions. *Journal of Human Evolution*, 36(5), 487–517. <https://doi.org/10.1006/jhev.1998.0287>
- Lieberman, D. E., Ross, C. F., & Ravosa, M. J. (2000). The primate cranial base: Ontogeny, function, and integration. *American Journal of Physical Anthropology*, 113(S31), 117–169. [https://doi.org/10.1002/1096-8644\(2000\)43:31+%3c117::AID-AJPA5%3e3.0.CO;2-I](https://doi.org/10.1002/1096-8644(2000)43:31+%3c117::AID-AJPA5%3e3.0.CO;2-I)
- Lieberman, D. E., McCarthy, R. C., Hiiemae, K. M., & Palmer, J. B. (2001). Ontogeny of postnatal hyoid and larynx descent in humans. *Archives of Oral Biology*, 46(2), 117–128. [https://doi.org/10.1016/S0003-9969\(00\)00108-4](https://doi.org/10.1016/S0003-9969(00)00108-4)
- Lieberman, D. E., McBratney, B. M., & Krovitz, G. (2002). The evolution and development of cranial form in Homo sapiens. *Proceedings of the National Academy of Sciences*, 99(3), 1134–1139. <https://doi.org/10.1073/pnas.022440799>
- Lieberman, D. (2011). *The evolution of the human head*. Harvard University Press.
- Lieberman, P. H., Klatt, D. H., & Wilson, W. H. (1969). Vocal tract limitations on the vowel repertoires of rhesus monkey and other nonhuman primates. *Science*, 164(3884), 1185–1187. <https://doi.org/10.1126/science.164.3884.1185>
- Lieberman, P., & Crelin, E. S. (1971). On the speech of Neanderthal man. *JANUA LINGUARUM*, 76.
- Lieberman, P., Crelin, E. S., & Klatt, D. H. (1972). Phonetic ability and related anatomy of the newborn and adult human, Neanderthal man, and the chimpanzee. *American Anthropologist*, 74(3), 287–307. <https://doi.org/10.1525/aa.1972.74.3.02a00020>
- Lieberman, P. (1984). *The biology and evolution of language*. Harvard University Press.
- Lieberman, P., Laitman, J. T., Reidenberg, J. S., & Gannon, P. J. (1992). The anatomy, physiology, acoustics and perception of speech: Essential elements in analysis of the evolution of human speech. *Journal of Human Evolution*, 23(6), 447–467. [https://doi.org/10.1016/0047-2484\(92\)90046-C](https://doi.org/10.1016/0047-2484(92)90046-C)
- Lieberman, P. (2000). *Human language and our reptilian brain: The subcortical bases of speech, syntax, and thought*. Harvard University Press.
- Lieberman, P. (2006). *Toward an evolutionary biology of language*. Harvard University Press.
- Lieberman, P. (2007). The evolution of human speech: Its anatomical and neural bases. *Current Anthropology*, 48(1), 39–66. <https://doi.org/10.1086/509092>
- Lieberman, P. (2012). Vocal tract anatomy and the neural bases of talking. *Journal of Phonetics*, 40(4), 608–622.
- Lieberman, P., McCarthy, R.C. (2015). The Evolution of Speech and Language. In W. Henke and I. Tattersall (Eds.), *Handbook of Paleoanthropology*. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-27800-6\\_79-1](https://doi.org/10.1007/978-3-642-27800-6_79-1)
- Lieberman, P. (2017). Comment on “Monkey vocal tracts are speech-ready.” *Science Advances*, 3(7), e1700442. <https://doi.org/10.1126/sciadv.1700442>
- Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, 839–862. <https://doi.org/10.2307/411991>
- Lindenfors, P., Wartel, A., & Lind, J. (2021). ‘Dunbar’s number’ deconstructed. *Biology Letters*, 17(5), 20210158. <https://doi.org/10.1098/rsbl.2021.0158>
- Lindblom, B. (1963). Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America*, 35(11), 1773–1781. <https://doi.org/10.1121/1.1918816>
- Lindblom, B. (1983). Economy of speech gestures. In P. MacNeilage (Ed.), *The production of speech*. Springer, 217–245. [https://doi.org/10.1007/978-1-4613-8202-7\\_10](https://doi.org/10.1007/978-1-4613-8202-7_10)
- Lindblom, B., MacNeilage, P., & Studdert-Kennedy, M. (1983). Self-organizing processes and the explanation of phonological universals. In B. Butterworth, B. Comrie, & Ö. Dahl (Eds.), *Explanations for language universals*. DeGruyter, 181–204. <https://doi.org/10.1515/ling.1983.21.1.181>

- Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle, & A. Marchal (Eds.), *Speech production and speech modelling*. Springer, 403–439. [https://doi.org/10.1007/978-94-009-2037-8\\_16](https://doi.org/10.1007/978-94-009-2037-8_16)
- Lindblom, B., Sussman, H. M., & Agwuele, A. (2009). A duration-dependent account of coarticulation for hyper- and hypoarticulation. *Phonetica*, 66(3), 188–195. <https://doi.org/10.1159/000235660>
- Lindblom, B., & MacNeilage, P. (2011). Coarticulation: A universal phonetic phenomenon with roots in deep time. *TMH-QPSR*, 51, 41–44.
- Liu, Z., Xu, Y., & Hsieh, F. F. (2022). Coarticulation as synchronised CV co-onset—Parallel evidence from articulation and acoustics. *Journal of Phonetics*, 90, 101116. <https://doi.org/10.1016/j.wocn.2021.101116>
- Lockwood, C. A. (1999). Sexual dimorphism in the face of *Australopithecus africanus*. *American Journal of Physical Anthropology*, 108(1), 97–127. [https://doi.org/10.1002/\(SICI\)1096-8644\(199901\)108:1%3c97::AID-AJPA6%3e3.0.CO;2-O](https://doi.org/10.1002/(SICI)1096-8644(199901)108:1%3c97::AID-AJPA6%3e3.0.CO;2-O)
- Lombard, M., & Gärdenfors, P. (2023). Minds on Fire: Cognitive Aspects of Early Firemaking and the Possible Inventors of Firemaking Kits. *Cambridge Archaeological Journal*, 1–21. <https://doi.org/10.1017/S0959774322000439>
- Lordkipanidze, D., Ponce de León, M. S., Margvelashvili, A., Rak, Y., Rightmire, G. P., Vekua, A., & Zollikofer, C. P. (2013). A complete skull from Dmanisi, Georgia, and the evolutionary biology of early Homo. *Science*, 342(6156), 326–331. <https://doi.org/10.1126/science.1238484>
- Lucas, P. W., Ang, K. Y., Sui, Z., Agrawal, K. R., Prinz, J. F., & Dominy, N. J. (2006). A brief review of the recent evolution of the human mouth in physiological and nutritional contexts. *Physiology & Behavior*, 89(1), 36–38. <https://doi.org/10.1016/j.physbeh.2006.03.016>
- Lupo, K. D., & Schmitt, D. N. (2005). Small prey hunting technology and zooarchaeological measures of taxonomic diversity and abundance: Ethnoarchaeological evidence from Central African forest foragers. *Journal of Anthropological Archaeology*, 24(4), 335–353. <https://doi.org/10.1016/j.jaa.2005.02.002>
- Logan, C. J., Avin, S., Boogert, N., Buskell, A., Cross, F. R., Currie, A., ... & Montgomery, S. H. (2018). Beyond brain size: Uncovering the neural correlates of behavioral and cognitive specialization. *Comparative Cognition & Behavior Reviews*, 13, 55–89. <https://doi.org/10.3819/CCBR.2018.130008>
- Lombard, M., & van Aardt, A. (2023). Method for generating foodplant fitness landscapes: With a foodplant checklist for southern Africa and its application to Klasies River Main Site. *Journal of Archaeological Science*, 149, 105707. <https://doi.org/10.1016/j.jas.2022.105707>
- Lund, J. P., & Kolta, A. (2006). Generation of the central masticatory pattern and its modification by sensory feedback. *Dysphagia*, 21(3), 167–174. <https://doi.org/10.1007/s00455-006-9027-6>
- MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, 21(4), 499–511. <https://doi.org/10.1017/S0140525X98001265>
- Mahajan, P. V., & Bharucha, B. A. (1994). Evaluation of short neck: New neck length percentiles and linear correlations with height and sitting height. *Indian Pediatrics*, 31(10), 1193–1203.
- Markov, A. V., & Markov, M. A. (2020). Runaway brain-culture coevolution as a reason for larger brains: Exploring the “cultural drive” hypothesis by computer modeling. *Ecology and Evolution*, 10(12), 6059–6077. <https://doi.org/10.1002/ece3.6350>
- Martin, R. D., Chivers, D. J., MacLarnon, A. M., & Hladik, C. M. (1985). Gastrointestinal allometry in primates and other mammals. In W. L. Jungers (Ed.), *Size and scaling in primate biology* (pp. 61–89). Springer.
- Marzke, M. W., & Shackley, M. S. (1986). Hominid hand use in the Pliocene and Pleistocene: Evidence from experimental archaeology and comparative morphology. *Journal of Human Evolution*, 15(6), 439–460. [https://doi.org/10.1016/S0047-2484\(86\)80027-6](https://doi.org/10.1016/S0047-2484(86)80027-6)
- Marzke, M. W. (1997). Precision grips, hand morphology, and tools. *American Journal of Physical Anthropology*, 102(1), 91–110. [https://doi.org/10.1002/\(SICI\)1096-8644\(199701\)102:1%3c91::AID-AJPA8%3e3.0.CO;2-G](https://doi.org/10.1002/(SICI)1096-8644(199701)102:1%3c91::AID-AJPA8%3e3.0.CO;2-G)
- Marzke, M. W. (2013). Tool making, hand morphology and fossil hominins. *Philosophical Transactions of the Royal Society b: Biological Sciences*, 368(1630), 20120414. <https://doi.org/10.1098/rstb.2012.0414>
- McCarthy, D. (1946). “Language development in children,” in L. Carmichael (ed.), *Manual of Child Psychology*. 2nd Ed. New York: John Wiley & Sons, Inc.
- McHenry, H. M. (1992). Body size and proportions in early hominids. *American Journal of Physical Anthropology*, 87(4), 407–431. <https://doi.org/10.1002/ajpa.1330870404>

- MacLarnon, A. M. (1987). *Size relationships of the spinal cord and associated skeleton in primates* [Doctoral dissertation]. University College London.
- MacLarnon, A., & Hewitt, G. (2004). Increased breathing control: Another factor in the evolution of human language. *Evolutionary Anthropology*, 13(5), 181–197. <https://doi.org/10.1002/evan.20032>
- McCauley, B., Collard, M., & Sandgathe, D. (2020). A cross-cultural survey of on-site fire use by recent hunter-gatherers: Implications for research on Palaeolithic pyrotechnology. *Journal of Paleolithic Archaeology*, 3, 566–584. <https://doi.org/10.1007/s41982-020-00052-7>
- McMurray, B. (2022). The myth of categorical perception. *The Journal of the Acoustical Society of America*, 152(6), 3819–3842. <https://doi.org/10.1121/10.0016614>
- McPherron, S. P., Alemseged, Z., Marean, C. W., Wynn, J. G., Reed, D., Geraads, D., ... & Béarat, H. A. (2010). Evidence for stone-tool-assisted consumption of animal tissues before 3.39 million years ago at Dikika, Ethiopia. *Nature*, 466(7308), 857–860. <https://doi.org/10.1038/nature09248>
- Meyer, M., Arsuaga, J. L., De Filippo, C., Nagel, S., Aximu-Petri, A., Nickel, B., ... & Pääbo, S. (2016). Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature*, 531(7595), 504–507. <https://doi.org/10.1038/nature17405>
- Metfessel, M. (1935). Roller canary song produced without learning from external sources. *Science*, 81(2106), 470–470. <https://doi.org/10.1126/science.81.2106.470.a>
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81–97. <https://doi.org/10.1037/h0043158>
- Milton, K. (1987). Primate diets and gut morphology: implications for hominid evolution. In M. Harris & E. B. Ross (Eds.), *Food and evolution: toward a theory of human food habits*. Temple University Press, 93–115.
- Milton, K. (1999). A hypothesis to explain the role of meat-eating in human evolution. *Evolutionary Anthropology: Issues News and Reviews*, 8(1), 11–21. [https://doi.org/10.1002/\(SICI\)1520-6505\(1999\)8:1%3c11::AID-EVAN6%3e3.0.CO;2-M](https://doi.org/10.1002/(SICI)1520-6505(1999)8:1%3c11::AID-EVAN6%3e3.0.CO;2-M)
- Montgomery, S. H. (2013). Primate brains, the ‘island rule’ and the evolution of *Homo floresiensis*. *Journal of Human Evolution*, 65(6), 750–760. <https://doi.org/10.1016/j.jhevol.2013.08.006>
- Montgomery, S. (2018). Hominin brain evolution: The only way is up? *Current Biology*, 28(14), R788–R790. <https://doi.org/10.1016/j.cub.2018.06.021>
- Morgan, A., Fisher, S. E., Scheffer, I., & Hildebrand, M. (2023). FOXP2-related speech and language disorder. In *GeneReviews®*. University of Washington, Seattle. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK368474/>
- Moran, S., & McCloy, D. (2019). PHOIBLE 2.0. *Jena: Max Planck Institute for the Science of Human History*.
- Moran, S., Lester, N. A., & Grossman, E. (2021). Inferring recent evolutionary changes in speech sounds. *Philosophical Transactions of the Royal Society B*, 376(1824), 20200198. <https://doi.org/10.1098/rstb.2020.0198>
- Moran, S., Kirkham, S., Friedrichs, D., Strait, D. S., & Ekström, A. G. (2024). Vocal tract proportions and the evolution of speech: New data to answer old questions. In M. Heldner, M. Włodarczak, M., C. Ericsdotter Nordgren., C. Wikse Barrow (Eds.), *Proceedings from FONETIK 2024*, 109–114. 10.5281/zenodo.11396092
- Motes-Rodrigo, A., & Tennie, C. (2021). The method of local restriction: In search of potential great ape culture-dependent forms. *Biological Reviews*, 96(4), 1441–1461. <https://doi.org/10.1111/brv.12710>
- Möttönen, R., & Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *Journal of Neuroscience*, 29(31), 9819–9825. <https://doi.org/10.1523/JNEUROSCI.6018-08.2009>
- Mulcahy, N. J., & Call, J. (2006). Apes save tools for future use. *Science*, 312(5776), 1038–1040. <https://doi.org/10.1126/science.1125456>
- Murdoch, B. E. (2001). Subcortical brain mechanisms in speech and language. *Folia Phoniatrica Et Logopaedica*, 53(5), 233–251. <https://doi.org/10.1159/000052679>
- Nearey, T. (1978). *Phonetic features for vowels*. Indiana University Linguistics Club.
- Negus, V. E. (1949). *The comparative anatomy and physiology of the larynx*. Heinemann.
- Neubauer, S., Gunz, P., Schwarz, U., Hublin, J. J., & Boesch, C. (2012). Brief communication: Endocranial volumes in an ontogenetic sample of chimpanzees from the Taï Forest National Park, Ivory Coast. *American Journal of Physical Anthropology*, 147(2), 319–325. <https://doi.org/10.1002/ajpa.21641>



- Nishimura, T. (2005). Developmental changes in the shape of the supralaryngeal vocal tract in chimpanzees. *American Journal of Physical Anthropology*, 126(2), 193–204. <https://doi.org/10.1002/ajpa.20112>
- Nishimura, T., Tokuda, I. T., Miyachi, S., Dunn, J. C., Herbst, C. T., Ishimura, K., ... & Fitch, W. T. (2022). Evolutionary loss of complexity in human vocal anatomy as an adaptation for speech. *Science*, 377(6607), 760–763. <https://doi.org/10.1126/science.abm1574>
- Öhman, S. E. (1967). Numerical model of coarticulation. *The Journal of the Acoustical Society of America*, 41(2), 310–320.
- Organ, C., Nunn, C. L., Machanda, Z., & Wrangham, R. W. (2011). Phylogenetic rate shifts in feeding time during the evolution of Homo. *Proceedings of the National Academy of Sciences*, 108(35), 14555–14559. <https://doi.org/10.1073/pnas.1107806108>
- Osvath, M., & Gärdenfors, P. (2005). Oldowan culture and the evolution of anticipatory cognition. *Lund University Cognitive Studies*, 122, 1–16. <https://doi.org/10.1121/1.1910340>
- Osvath, M. (2009). Spontaneous planning for future stone throwing by a male chimpanzee. *Current Biology*, 19(5), R190–R191. <https://doi.org/10.1016/j.cub.2009.01.010>
- Osvath, M., & Osvath, H. (2008). Chimpanzee (*Pan troglodytes*) and orangutan (*Pongo abelii*) forethought: Self-control and pre-experience in the face of future tool use. *Animal Cognition*, 11, 661–674. <https://doi.org/10.1007/s10071-008-0157-0>
- Pal, A., Kumara, H. N., Mishra, P. S., Velankar, A. D., & Singh, M. (2018). Extractive foraging and tool-aided behaviors in the wild Nicobar long-tailed macaque (*Macaca fascicularis umbrosus*). *Primates*, 59(2), 173–183. <https://doi.org/10.1007/s10329-017-0635-6>
- Palmer, J. B., Rudin, N. J., Lara, G., & Crompton, A. W. (1992). Coordination of mastication and swallowing. *Dysphagia*, 7(4), 187–200. <https://doi.org/10.1007/BF02493469>
- Panger, M. A., Brooks, A. S., Richmond, B. G., & Wood, B. (2002). Older than the Oldowan? Rethinking the emergence of hominin tool use. *Evolutionary Anthropology: Issues, News, and Reviews*, 11(6), 235–245. <https://doi.org/10.1002/evan.10094>
- Pante, M. C. (2013). The larger mammal fossil assemblage from JK2, Bed III, Olduvai Gorge, Tanzania: Implications for the feeding behavior of Homo erectus. *Journal of Human Evolution*, 64(1), 68–82. <https://doi.org/10.1016/j.jhevol.2012.10.004>
- Pante, M. C., Njau, J. K., Hensley-Marschand, B., Keevil, T. L., Martín-Ramos, C., Peters, R. F., & de la Torre, I. (2018). The carnivorous feeding behavior of early Homo at HWK EE, Bed II, Olduvai Gorge, Tanzania. *Journal of Human Evolution*, 120, 215–235. <https://doi.org/10.1016/j.jhevol.2017.06.005>
- Perlman, M., Patterson, F. G., & Cohn, R. H. (2012). The human-fostered gorilla Koko shows breath control in play with wind instruments. *Biolinguistics*, 6(3–4), 433–444. <https://doi.org/10.5964/bioling.8935>
- Perlman, M., & Clark, N. (2015). Learned vocal and breathing behavior in an enculturated gorilla. *Animal Cognition*, 18, 1165–1179. <https://doi.org/10.1007/s10071-015-0889-6>
- Pereira, A. S., Kavanagh, E., Hobaiter, C., Slocombe, K. E., & Lameira, A. R. (2020). Chimpanzee lip-smacks confirm primate continuity for speech-rhythm evolution. *Biology Letters*, 16(5), 20200232. <https://doi.org/10.1098/rsbl.2020.0232>
- Peters, C. R. (1987). Nut-like oil seeds: Food for monkeys, chimpanzees, humans, and probably ape-men. *American Journal of Physical Anthropology*, 73(3), 333–363. <https://doi.org/10.1002/ajpa.1330730306>
- Peters, B., Bledowski, C., Rieder, M., & Kaiser, J. (2016). Recurrence of task set-related MEG signal patterns during auditory working memory. *Brain Research*, 1640, 232–242. <https://doi.org/10.1016/j.brainres.2015.12.006>
- Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, 24(2), 175–184. <https://doi.org/10.1121/1.1906875>
- Pidoux, L., Le Blanc, P., Levenes, C., & Leblois, A. (2018). A subcortical circuit linking the cerebellum to the basal ganglia engaged in vocal learning. *eLife*, 7, e32167. <https://doi.org/10.7554/eLife.32167>
- Piette, T., Cathcart, C., Babrieri, C., Grandjean, D., Déaux, É., & Giraud, A.-L. (2022). Theta rhythm is widespread in vocal production across the animal realm. In A. Ravignani, R. Asano, D. Valente, F. Ferretti, S. Hartmann, M. Hayashi, Y. Jadoul, M. Martins, Y. Oseki, E. D. Rodrigues, O. Vasileva, & S. Waciewicz (Eds.), *Proceedings of the Joint Conference on Language Evolution (JCoLE)* (pp. 579–582). Joint Conference on Language Evolution (JCoLE); Max Planck Institute for Psycholinguistics. [http://www.evolang.org/jcole\\_proceedings/jcole\\_proceedings.pdf](http://www.evolang.org/jcole_proceedings/jcole_proceedings.pdf)

- Pisoni, D. B. (1971). *On the nature of categorical perception of speech sounds*. University of Michigan.
- Planer, R., & Sterelny, K. (2021). *From signal to symbol: The evolution of language*. MIT Press.
- Plavcan, J. M., & van Schaik, C. P. (1992). Intrasexual competition and canine dimorphism 891 in anthropoid primates. *American Journal of Physical Anthropology*, 87(4), 461–477. <https://doi.org/10.1002/ajpa.1330870407>
- Plavcan, J. M., & Van Schaik, C. P. (1997). Interpreting hominid behavior on the basis of sexual dimorphism. *Journal of Human Evolution*, 32(4), 345–374. <https://doi.org/10.1006/jhev.1996.0096>
- Plavcan, J. M. (2001). Sexual dimorphism in primate evolution. *American Journal of Physical Anthropology*, 116(S33), 25–53. <https://doi.org/10.1002/ajpa.10011>
- Plummer, T. W., Oliver, J. S., Finestone, E. M., Ditchfield, P. W., Bishop, L. C., Blumenthal, S. A., Lemorini, C., ..., & Potts, R. (2023). Expanded geographic distribution and dietary strategies of the earliest Oldowan hominins and Paranthropus. *Science*, 379(6632), 561–566. <https://doi.org/10.1126/science.abo7452>
- Poeppel, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature Reviews Neuroscience*, 21(6), 322–334. <https://doi.org/10.1038/s41583-020-0304-4>
- Ponce de León, M. S., Bienvenu, T., Marom, A., Engel, S., Tafforeau, P., Alatorre Warren, J. L., ... & Zollikofer, C. P. (2021). The primitive brain of early Homo. *Science*, 372(6538), 165–171. <https://doi.org/10.1126/science.aaz0032>
- Pruetz, J. D., & Herzog, N. M. (2017). Savanna chimpanzees at Fongoli, Senegal, navigate a fire landscape. *Current Anthropology*, 58(S16), S337–S350. <https://doi.org/10.1086/692112>
- Pruetz, J. D., & LaDuke, T. C. (2010). Brief communication: Reaction to fire by savanna chimpanzees (*Pan troglodytes verus*) at Fongoli, Senegal: Conceptualization of "fire behavior" and the case for a chimpanzee model. *American Journal of Physical Anthropology*, 141(4), 646–650. <https://doi.org/10.1002/ajpa.21245>
- Pu, L., Fang, C., Hsing-Hua, M., Ching-Yu, P., Li-Sheng, H., & Shih-Chiang, C. (1977). Preliminary study on the age of Yuanmou man by palaeomagnetic technique. *Scientia Sinica*, 20(5), 645–664.
- Puts, D. A., Jones, B. C., & DeBruine, L. M. (2012). Sexual selection on human faces and voices. *Journal of Sex Research*, 49(2-3), 227–243. Puts, D. A., Jones, B. C., & DeBruine, L. M. (2012). Sexual selection on human faces and voices. *Journal of Sex Research*, 49(2-3), 227–243
- Ragir, S. (2000). Diet and food preparation: Rethinking early hominid behavior. *Evolutionary Anthropology: Issues, News, and Reviews*, 9(4), 153–155. [https://doi.org/10.1002/1520-6505\(2000\)9:4%3c153::AID-EVAN4%3e3.0.CO;2-D](https://doi.org/10.1002/1520-6505(2000)9:4%3c153::AID-EVAN4%3e3.0.CO;2-D)
- Ragir, S., Rosenberg, M., & Tierno, P. (2000). Gut morphology and the avoidance of carrion among chimpanzees, baboons, and early hominids. *Journal of Anthropological Research*, 56(4), 477–512. <https://doi.org/10.1086/jar.56.4.3630928>
- Rightmire, G. P. (2013). Homo erectus and Middle Pleistocene hominins: Brain size, skull form, and species recognition. *Journal of Human Evolution*, 65(3), 223–252. <https://doi.org/10.1016/j.jhev.2013.04.008>
- Roebroeks, W., & Villa, P. (2011). On the earliest evidence for habitual use of fire in Europe. *Proceedings of the National Academy of Sciences*, 108(13), 5209–5214. <https://doi.org/10.1073/pnas.1018116108>
- Rose, L., & Marshall, F. (1996). Meat eating, hominid sociality, and home bases revisited. *Current Anthropology*, 37(2), 307–338. <https://doi.org/10.1086/204494>
- Ross, C. F., Washington, R. L., Eckhardt, A., Reed, D. A., Vogel, E. R., Dominy, N. J., & Machanda, Z. P. (2009). Ecological consequences of scaling of chew cycle duration and daily feeding time in Primates. *Journal of Human Evolution*, 56(6), 570–585. <https://doi.org/10.1016/j.jhev.2009.02.007>
- Russell, J. L., McIntyre, J. M., Hopkins, W. D., & Tagliatela, J. P. (2013). Vocal learning of a communicative signal in captive chimpanzees. *Pan Troglodytes. Brain and Language*, 127(3), 520–525. <https://doi.org/10.1016/j.bandl.2013.09.009>
- Sagum, R., & Arcot, J. (2000). Effect of domestic processing methods on the starch, non-starch polysaccharides and in vitro starch and protein digestibility of three varieties of rice with varying levels of amylose. *Food Chemistry*, 70(1), 107–111. [https://doi.org/10.1016/S0308-8146\(00\)00041-8](https://doi.org/10.1016/S0308-8146(00)00041-8)
- Salmi, R., Szczupider, M., & Carrigan, J. (2022a). A novel attention-getting vocalization in zoo-housed western gorillas. *PLoS ONE*, 17(8), e0271871. <https://doi.org/10.1371/journal.pone.0271871>
- Salmi, R., Jones, C. E., & Carrigan, J. (2022b). Who is there? Captive western gorillas distinguish human voices based on familiarity and nature of previous interactions. *Animal Cognition*, 25(1), 217–228. <https://doi.org/10.1007/s10071-021-01543-y>

- Sanz, C., Call, J., & Morgan, D. (2009). Design complexity in termite-fishing tools of chimpanzees (*Pan troglodytes*). *Biology Letters*, 5(3), 293–296. <https://doi.org/10.1098/rsbl.2008.0786>
- Sato, K., Nishimura, T., Sato, K., Sato, F., Chitose, S. I., & Umeno, H. (2023). Comparative Histoanatomy of the Epiglottis and Pre-epiglottic Space of the Chimpanzee Larynx. *Journal of Voice*. <https://doi.org/10.1016/j.jvoice.2023.07.027>
- Schick, K. D., & Toth, N. P. (1994). *Making silent stones speak: Human evolution and the dawn of technology*. Simon and Schuster.
- Schwob, N. (2017). Evidence of language prerequisites in Pan: orofacial-motor and breath control in chimpanzees and bonobos. [Master's thesis.] Kennesaw State University.
- Semaw, S., Renne, P., Harris, J. W., Feibel, C. S., Bernor, R. L., Fesseha, N., & Mowbray, K. (1997). 2.5-million-year-old stone tools from Gona, Ethiopia. *Nature*, 385(6614), 333–336. <https://doi.org/10.1038/385333a0>
- Semaw, S., Rogers, M. J., Quade, J., Renne, P. R., Butler, R. F., Dominguez-Rodrigo, M., Stout, D., Hart, W. S., Pickering, T., & Simpson, S. W. (2003). 2.6-Million-year-old stone tools and associated bones from OGS-6 and OGS-7, Gona, Afar, Ethiopia. *Journal of Human Evolution*, 45(2), 169–177. [https://doi.org/10.1016/S0047-2484\(03\)00093-9](https://doi.org/10.1016/S0047-2484(03)00093-9)
- Seyfarth, R. M., & Cheney, D. L. (2014). The evolution of language from social cognition. *Current Opinion in Neurobiology*, 28, 5–9. <https://doi.org/10.1016/j.conb.2014.04.003>
- Shiffrin, R. M., & Nosofsky, R. M. (1994). Seven plus or minus two: A commentary on capacity limitations. *Psychological Review*, 101(2), 357–361. <https://doi.org/10.1037/0033-295X.101.2.357>
- Shimelmitz, R., Kuhn, S. L., Jelinek, A. J., Ronen, A., Clark, A. E., & Weinstein-Evron, M. (2014). 'Fire at will': The emergence of habitual fire use 350,000 years ago. *Journal of Human Evolution*, 77, 196–203. <https://doi.org/10.1016/j.jhevol.2014.07.005>
- Shipley, C., Carterette, E. C., & Buchwald, J. S. (1991). The effects of articulation on the acoustical structure of feline vocalizations. *The Journal of the Acoustical Society of America*, 89(2), 902–909. <https://doi.org/10.1121/1.1894652>
- Schön Ybarra, M. A. (1995). A comparative approach to the non-human primate vocal tract: Implications for sound production. In E. Zimmerman, J. D. Newman & U. Jürgens (Eds.), *Current topics in primate vocal communication*. Springer, 185–198. [https://doi.org/10.1007/978-1-4757-9930-9\\_9](https://doi.org/10.1007/978-1-4757-9930-9_9)
- Schötz, S. (2020). Phonetic Variation in Cat–Human Communication. In M. Ramiro Pastorinho & A. C. A. Soursa (Eds.), *Pets as Sentinels, Forecasters and Promoters of Human Health*. Springer, 319–347. [https://doi.org/10.1007/978-3-030-30734-9\\_14](https://doi.org/10.1007/978-3-030-30734-9_14)
- Shultz, S., Nelson, E., & Dunbar, R. I. (2012). Hominin cognitive evolution: Identifying patterns and processes in the fossil and archaeological record. *Philosophical Transactions of the Royal Society b: Biological Sciences*, 367(1599), 2130–2140. <https://doi.org/10.1098/rstb.2012.0115>
- Shofner, W. P. (2024). What's special about human speech? A student exercise for comparing speech production between humans and chimpanzees. *The Journal of the Acoustical Society of America*, 155(5), 3206–3212. <https://doi.org/10.1121/10.0026020>
- Smith, C. S., Martin, W., & Johansen, K. A. (2001). Sego lilies and prehistoric foragers: Return rates, pit ovens, and carbohydrates. *Journal of Archaeological Science*, 28(2), 169–183. <https://doi.org/10.1006/jasc.2000.0554>
- Smith, A. R., Carmody, R. N., Dutton, R. J., & Wrangham, R. W. (2015a). The significance of cooking for early hominin scavenging. *Journal of Human Evolution*, 84, 62–70. <https://doi.org/10.1016/j.jhevol.2015.03.013>
- Smith, A. L., Benazzi, S., Ledogar, J. A., Tamvada, K., Pryor Smith, L. C., Weber, G. W., ... & Strait, D. S. (2015). The feeding biomechanics and dietary ecology of *Paranthropus boisei*. *The Anatomical Record*, 298(1), 145–167. <https://doi.org/10.1002/ar.23073>
- Snyder, W. D., Reeves, J. S., & Tennie, C. (2022). Early knapping techniques do not necessitate cultural transmission. *Science Advances*, 8(27), eabo2894. <https://doi.org/10.1126/sciadv.abo2894>
- Speth, J. D. (2017). Putrid meat and fish in the Eurasian middle and upper paleolithic: Are we missing key part of Neanderthal and modern human diet? *PaleoAnthropology*, 44–72. <https://doi.org/10.4207/PA.2017.ART105>
- Stahl, A. B., Dunbar, R. I. M., Homewood, K., Ikawa-Smith, F., Kortlandt, A., McGrew, ... & Wrangham, R. W. (1984). Hominid diet before fire. *Current Anthropology*, 25, 151–168. <https://doi.org/10.1086/203106>

- Stedman, H. H., Kozyak, B. W., Nelson, A., Thesier, D. M., Su, L. T., Low, D. W., ... & Mitchell, M. A. (2004). Myosin gene mutation correlates with anatomical changes in the human lineage. *Nature*, 428(6981), 415–418. <https://doi.org/10.1038/nature02358>
- Steele, J., Clegg, M., & Martelli, S. (2013). Comparative morphology of the hominin and african ape hyoid bone, a possible marker of the evolution of speech. *Human Biology*, 85(5), 639–672. <https://doi.org/10.3378/027.085.0501>
- Stevens, K. N. (1969). Evidence for quantal vowel articulations. *The Journal of the Acoustical Society of America*, 46(1A\_Supplement), 110–110. <https://doi.org/10.1121/1.1972528>
- Stevens, K. N. (1972). The quantal nature of speech: evidence from articulatory-acoustic data. In P. B. Dennes and E. E. David Jr (Eds.), *Human Communication, A Unified View*. New York: McGraw-Hill.
- Stevens, K. N., & Blumstein, S. E. (1975). Quantal aspects of consonant production and perception: A study of retroflex consonants. *Journal of Phonetics*, 3, 215–233. [https://doi.org/10.1016/S0095-4470\(19\)31431-7](https://doi.org/10.1016/S0095-4470(19)31431-7)
- Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, 17(1), 3–45.
- Stoinski, T. S., & Whiten, A. (2003). Social learning by orangutans (*Pongo abelii* and *Pongo pygmaeus*) in a simulated food-processing task. *Journal of Comparative Psychology*, 117(3), 272. <https://doi.org/10.1037/0735-7036.117.3.272>
- Strait, D. S., & Grine, F. E. (2004). Inferring hominoid and early hominid phylogeny using craniodental characters: The role of fossil taxa. *Journal of Human Evolution*, 47(6), 399–452. <https://doi.org/10.1016/j.jhev.2004.08.008>
- Strait, D. S., Weber, G. W., Neubauer, S., Chalk, J., Richmond, B. G., Lucas, P. W., ... & Smith, A. L. (2009). The feeding biomechanics and dietary ecology of *Australopithecus africanus*. *Proceedings of the National Academy of Sciences*, 106(7), 2124–2129. <https://doi.org/10.1073/pnas.0808730106>
- Strait, D. S., Constantino, P., Lucas, P. W., Richmond, B. G., Spencer, M. A., Dechow, P. C., ... & Ledogar, J. A. (2013). Viewpoints: diet and dietary adaptations in early hominins: the hard food perspective. *American Journal of Physical Anthropology*, 151(3), 339–355. <https://doi.org/10.1002/ajpa.22285>
- Studdert-Kennedy, M. (1998). The particulate origins of language generativity: from syllable to gesture. In (Hurford, J. et al., Eds.), *Approaches to the Evolution of Language: Social and Cognitive Bases*, pp. 202–221, Cambridge University Press.
- Susman, R. L. (1994). Fossil evidence for early hominid tool use. *Science*, 265(5178), 1570–1573. <https://doi.org/10.1126/science.8079169>
- Susman, R. L. (1998). Hand function and tool behavior in early hominids. *Journal of Human Evolution*, 35(1), 23–46. <https://doi.org/10.1006/jhev.1998.0220>
- Sussman, E. S., & Gumenyuk, V. (2005). Organization of sequential sounds in auditory memory. *NeuroReport*, 16(13), 1519–1523. <https://doi.org/10.1097/01.wnr.0000177002.35193.4c>
- Taglialatela, J. P., Savage-Rumbaugh, S., & Baker, L. A. (2003). Vocal production by a language-competent *Pan paniscus*. *International Journal of Primatology*, 24, 1–17. <https://doi.org/10.1023/A:1021487710547>
- Taglialatela, J. P., Russell, J. L., Schaeffer, J. A., & Hopkins, W. D. (2008). Communicative signaling activates ‘Broca’s’ homolog in chimpanzees. *Current Biology*, 18(5), 343–348. <https://doi.org/10.1016/j.cub.2008.01.049>
- Taglialatela, J. P., Russell, J. L., Schaeffer, J. A., & Hopkins, W. D. (2009). Visualizing vocal perception in the chimpanzee brain. *Cerebral Cortex*, 19(5), 1151–1157. <https://doi.org/10.1093/cercor/bhn157>
- Taglialatela, J. P., Reamer, L., Schapiro, S. J., & Hopkins, W. D. (2012). Social learning of a communicative signal in captive chimpanzees. *Biology Letters*, 8(4), 498–501. <https://doi.org/10.1098/rsbl.2012.0113>
- Takemoto, H. (2008). Morphological analyses and 3D modeling of the tongue musculature of the chimpanzee (*Pan troglodytes*). *American Journal of Primatology*, 70(10), 966–975. <https://doi.org/10.1002/ajp.20589>
- Tan, A. W., Luncz, L., Haslam, M., Malaivijitnond, S., & Gumert, M. D. (2016). Complex processing of prickly pear cactus (*Opuntia* sp.) by free-ranging long-tailed macaques: preliminary analysis for hierarchical organisation. *Primates*, 57(2), 141–147. <https://doi.org/10.1007/s10329-016-0525-3>

- Thelen, E. (1991). Motor aspects of emergent speech: A dynamic approach. In N. Krasnegor, D. Rumbaugh, & M. Studdert-Kennedy (Eds.), *Biological and Behavioral Determinants of Language Development* (pp. 221–248). Academic Press.
- Thomas, J., & Kirby, S. (2018). Self domestication and the evolution of language. *Biology & Philosophy*, 33, 1–30. <https://doi.org/10.1007/s10539-018-9612-8>
- Tennie, C., Gilby, I. C., & Mundry, R. (2009). The meat-scrap hypothesis: Small quantities of meat may promote cooperative hunting in wild chimpanzees (*Pan troglodytes*). *Behavioral Ecology and Sociobiology*, 63, 421–431. <https://doi.org/10.1007/s00265-008-0676-3>
- Tennie, C., Premo, L. S., Braun, D. R., & McPherron, S. P. (2017). Early stone tools and cultural transmission: Resetting the null hypothesis. *Current Anthropology*, 58(5), 652–672. <https://doi.org/10.1086/693846>
- Tennie, C. (2023). The earliest tools and cultures of hominins. In J. J. Tehrani, J. Kendal, & R. Kendal (Eds.), *The Oxford handbook of cultural evolution*. <https://doi.org/10.1093/oxfordhb/9780198869252.013.33>
- Tobias, P. V. (1968). Cranial capacity in anthropoid apes, *Australopithecus* and *Homo habilis*, with comments on skewed samples. *South African Journal of Science*, 64(2), 81–91.
- Tobias, P. V. (1971). Human skeletal remains from the Cave of Hearths, Makapansgat, northern Transvaal. *American Journal of Physical Anthropology*, 34(3), 335–367. <https://doi.org/10.1002/ajpa.1330340305>
- Tomasello, M., Melis, A. P., Tennie, C., Wyman, E., & Herrmann, E. (2012). Two key steps in the evolution of human cooperation: The interdependence hypothesis. *Current Anthropology*, 53(6), 673–692. <https://doi.org/10.1086/668207>
- Toth, N., & Schick, K. (2018). An overview of the cognitive implications of the Oldowan Industrial Complex. *Azania: Archaeological Research in Africa*, 53(1), 3–39. <https://doi.org/10.1080/0067270X.2018.1439558>
- Trinkaus, E. (2003). Neandertal faces were not long; modern human faces are short. *Proceedings of the National Academy of Sciences*, 100(14), 8142–8145. <https://doi.org/10.1073/pnas.1433023100>
- Trinkaus, E. (2007). Human evolution: Neandertal gene speaks out. *Current Biology*, 17(21), R917–R919. <https://doi.org/10.1016/j.cub.2007.09.055>
- Vaesen, K. (2012). The cognitive bases of human tool use. *Behavioral and Brain Sciences*, 35(4), 203–218. <https://doi.org/10.1017/S0140525X11001452>
- Van Casteren, A., Codd, J. R., Kupczik, K., Plasqui, G., Sellers, W. I., & Henry, A. G. (2022). The cost of chewing: The energetics and evolutionary significance of mastication in humans. *Science Advances*, 8(33), eabn8351. <https://doi.org/10.1126/sciadv.abn8351>
- van Horik, J., & Emery, N. J. (2011). Evolution of cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(6), 621–633. <https://doi.org/10.1002/wcs.144>
- Vihman, M. M. (2014). *Phonological development: The first two years*. Wiley.
- Völter, C. J., & Call, J. (2014). The cognitive underpinnings of flexible tool use in great apes. *Journal of Experimental Psychology: Animal Learning and Cognition*, 40(3), 287. <https://doi.org/10.1037/xan0000025>
- von Cramon-Taubadel, N. (2017). Measuring the effects of farming on human skull morphology. *Proceedings of the National Academy of Sciences*, 114(34), 8917–8919. <https://doi.org/10.1073/pnas.1711475114>
- Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., & Yandell, B. S. (2005). Development of vocal tract length during early childhood: A magnetic resonance imaging study. *The Journal of the Acoustical Society of America*, 117(1), 338–350. <https://doi.org/10.1121/1.1835958>
- Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D., ... & Gentry, L. R. (2009). Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study. *The Journal of the Acoustical Society of America*, 125(3), 1666–1678. <https://doi.org/10.1121/1.3075589>
- Vekua, A., Lordkipanidze, D., Rightmire, G. P., Agusti, J., Ferring, R., Maisuradze, G., ... & Zollikofer, C. (2002). A new skull of early Homo from Dmanisi, Georgia. *Science*, 297(5578), 85–89. <https://doi.org/10.1126/science.1072953>
- Walker, M. J., Haber Uriarte, M., López Jiménez, A., López Martínez, M., Martín Lerma, I., Van der Made, J., ... & Grün, R. (2020). Cueva Negra del Estrecho del Río Quípar: A dated late early pleistocene palaeolithic site in Southeastern Spain. *Journal of Paleolithic Archaeology*, 3(4), 816–855. <https://doi.org/10.1007/s41982-020-00062-5>



- Warneken, F., & Rosati, A. G. (2015). Cognitive capacities for cooking in chimpanzees. *Proceedings of the Royal Society b: Biological Sciences*, 282(1809), 20150229. <https://doi.org/10.1098/rspb.2015.0229>
- Weissengruber, G. E., Forstenpointner, G., Peters, G., Kübber-Heiss, A., & 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*). *Journal of Anatomy*, 201(3), 195–209. <https://doi.org/10.1046/j.1469-7580.2002.00088.x>
- Whitehead, H. (2000). *Food rules: Hunting, sharing, and tabooing game in Papua New Guinea*. University of Michigan Press.
- White, T. D., Asfaw, B., DeGusta, D., Gilbert, H., Richards, G. D., Suwa, G., & Clark Howell, F. (2003). Pleistocene *Homo sapiens* from middle awash, ethiopia. *Nature*, 423(6941), 742–747.
- Wilsch, A., & Obleser, J. (2016). What works in auditory working memory? A neural oscillations perspective. *Brain Research*, 1640, 193–207. <https://doi.org/10.1016/j.brainres.2015.10.054>
- Wobber, V., Hare, B., & Wrangham, R. (2008). Great apes prefer cooked food. *Journal of Human Evolution*, 55(2), 340–348. <https://doi.org/10.1016/j.jhevol.2008.03.003>
- Wich, S. A., Swartz, K. B., Hardus, M. E., Lameira, A. R., Stromberg, E., & Shumaker, R. W. (2009). A case of spontaneous acquisition of a human sound by an orangutan. *Primates*, 50, 56–64. <https://doi.org/10.1007/s10329-008-0117-y>
- Wich, S. A., Krützen, M., Lameira, A. R., Nater, A., Arora, N., Bastian, M. L., ... & van Schaik, C. P. (2012). Call cultures in orang-utans?. *PLoS One*, 7(5), e36180. <https://doi.org/10.1371/journal.pone.0036180>
- Wood, S. (1979). A radiographic analysis of constriction locations for vowels. *Journal of Phonetics*, 7(1), 25–43. [https://doi.org/10.1016/S0095-4470\(19\)31031-9](https://doi.org/10.1016/S0095-4470(19)31031-9)
- Wood, S. (1986). The acoustical significance of tongue, lip, and larynx maneuvers in rounded palatal vowels. *The Journal of the Acoustical Society of America*, 80(2), 391–401. <https://doi.org/10.1121/1.394090>
- Wrangham, R. W. (1975). *Behavioural ecology of chimpanzees in Gombe National Park, Tanzania* [Doctoral dissertation]. University of Cambridge.
- Wrangham, R. W. (1977). Feeding behaviour of chimpanzees in Gombe National Park, Tanzania. In (T. H. Clutton-Brock, Ed.) *Primate Ecology: Studies of Feeding and Ranging Behaviour in Lemurs, Monkeys and Apes*, pp. 503–538. London: Academic Press.
- Wrangham, R. W., Jones, J. H., Laden, G., Pilbeam, D., & Conklin-Brittain, N. (1999). The raw and the stolen: Cooking and the ecology of human origins. *Current Anthropology*, 40(5), 567–594. <https://doi.org/10.1086/300083>
- Wrangham, R., & Conklin-Brittain, N. (2003). Cooking as a biological trait. *Comparative Biochemistry and Physiology Part a: Molecular & Integrative Physiology*, 136(1), 35–46. [https://doi.org/10.1016/S1095-6433\(03\)00020-5](https://doi.org/10.1016/S1095-6433(03)00020-5)
- Wrangham, R. (2009). *Catching fire: how cooking made us human*. Basic books.
- Wrangham, R. (2017). Control of fire in the Paleolithic: Evaluating the cooking hypothesis. *Current Anthropology*, 58(S16), S303–S313. <https://doi.org/10.1086/692113>
- Wrangham, R. W. (2019). Hypotheses for the evolution of reduced reactive aggression in the context of human self-domestication. *Frontiers in Psychology*, 10, 1914. <https://doi.org/10.3389/fpsyg.2019.01914>
- Yellen, J. E. (1991). Small mammals: !Kung San utilization and the production of faunal assemblages. *Journal of Anthropological Archaeology*, 10(1), 1–26. [https://doi.org/10.1016/0278-4165\(91\)90019-T](https://doi.org/10.1016/0278-4165(91)90019-T)
- Zeberg, H., Jakobsson, M., & Pääbo, S. (2024). The genetic changes that shaped Neandertals, Denisovans, and modern humans. *Cell*, 187(5), 1047–1058. <https://doi.org/10.1016/j.cell.2023.12.029>
- Zihlman, A. L., Stahl, D., & Boesch, C. (2008). Morphological variation in adult chimpanzees (*Pan troglodytes verus*) of the Taï National Park, Côte d'Ivoire. *American Journal of Physical Anthropology*, 135(1), 34–41. <https://doi.org/10.1002/ajpa.20702>
- Zink, K. D., Lieberman, D. E., & Lucas, P. W. (2014). Food material properties and early hominin processing techniques. *Journal of Human Evolution*, 77, 155–166. <https://doi.org/10.1016/j.jhevol.2014.06.012>
- Zink, K. D., & Lieberman, D. E. (2016). Impact of meat and Lower Palaeolithic food processing techniques on chewing in humans. *Nature*, 531(7595), 500–503. <https://doi.org/10.1038/nature16990>

- Zohar, I., Alperson-Afil, N., Goren-Inbar, N., Prévost, M., Tütken, T., Sisma-Ventura, G., ... & Najorka, J. (2022). Evidence for the cooking of fish 780,000 years ago at Gesher Benot Ya'aqov, Israel. *Nature Ecology & Evolution*, 6(12), 2016–2028. <https://doi.org/10.1038/s41559-022-01910-z>
- Zollikofer, C. P., Bienvenu, T., Beyene, Y., Suwa, G., Asfaw, B., White, T. D., & Ponce de León, M. S. (2022). Endocranial ontogeny and evolution in early *Homo sapiens*: The evidence from Herto, Ethiopia. *Proceedings of the National Academy of Sciences*, 119(32), e2123553119. <https://doi.org/10.1073/pnas.2123553119>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Axel G. Ekström** is Post-doctoral Researcher with the Institute of Biology at the University of Neuchâtel. His work is devoted to the study of comparative vocal anatomy, speech acoustics, and vocal tract modeling.

**Peter Gärdenfors** is Professor Emeritus at Lund University Cognitive Science, Lund University. His research is concentrated on concept formation, cognitive science, and the evolution of thinking.

**William D. Snyder** is a postdoctoral researcher associated with the working groups for Paleoanthropology and Early Prehistory and Quaternary Ecology at the University of Tübingen. His work combines various methodological and theoretical approaches for reconstructing prehistoric behavior of hominins and other primates.

**Daniel Friedrichs** is Senior Researcher with the Linguistic Research Infrastructure (LiRI) at the University of Zurich. His work deals with dynamics of speech production and perception in human speakers and listeners.

**Robert C. McCarthy** is Associate Professor at Benedictine University. His research is devoted to the study of the evolution and function of hominin craniofacial and vocal tract morphology.

**Melina Tsapos** is a PhD Student with the Department of Philosophy, Lund University. Her research is devoted to study and modeling of social cognition and sociocognitive processes.

**Claudio Tennie** is a permanent research group leader with the Early Prehistory & Quaternary Ecology Group at the University of Tübingen. His research spans across behavioural and evolutionary biology, comparative psychology, evolutionary archaeology and philosophy of cultural evolution, and focuses on the evolution of human cultural evolution.


**David S. Strait** is Professor of Anthropology at Washington University in St. Louis. His research focuses on the fossil record of human evolution, particularly the diversification of early hominins, the evolution of diet and feeding biomechanics.

**Jens Edlund** is Professor at the Division of Speech, Music & Hearing, at the KTH Royal Institute of Technology. His work is focused on studies of speech technology and cognitive science.

**Steven Moran** is Scientific Director of the Linguistic Research Infrastructure (LiRI) at the University of Zurich. His research is devoted to the study of phonology across languages and the evolution of speech capacities.



## Authors and Affiliations

**Axel G. Ekström<sup>1,2</sup>**  · **Peter Gärdenfors<sup>3,4</sup>** · **William D. Snyder<sup>5,6</sup>** · **Daniel Friedrichs<sup>2,7</sup>** · **Robert C. McCarthy<sup>8</sup>** · **Melina Tsapos<sup>3</sup>** · **Claudio Tennie<sup>6</sup>** · **David S. Strait<sup>4,9,10</sup>** · **Jens Edlund<sup>1</sup>** · **Steven Moran<sup>2,7,11</sup>**

✉ Axel G. Ekström  
axeleks@kth.se

Steven Moran  
steven.moran@uzh.ch

<sup>1</sup> Speech, Music & Hearing, KTH Royal Institute of Technology, Stockholm, Sweden

<sup>2</sup> Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland

<sup>3</sup> Department of Philosophy, Lund University, Lund, Sweden

<sup>4</sup> Paleo-Research Institute, University of Johannesburg, Johannesburg, South Africa

<sup>5</sup> Senckenberg Centre for Human Evolution and Palaeoenvironment, University of Tübingen, Tübingen, Germany

<sup>6</sup> Early Prehistory and Quaternary Ecology, Department of Geosciences, University of Tübingen, Tübingen, Germany

<sup>7</sup> Linguistics Research Infrastructure (LiRI), University of Zurich, Zürich, Switzerland

<sup>8</sup> Department of Biological Sciences, Benedictine University, Lisle, IL, US

<sup>9</sup> Department of Anthropology, Washington University in St. Louis, St. Louis, MO, US

<sup>10</sup> DFG Center for Advanced Studies “Words, Bones, Genes, Tools”, University of Tübingen, Tübingen, Germany

<sup>11</sup> Department of Anthropology, University of Miami, Coral Gables, FL, US