

A Theory That Never Was: Wrong Way to the “Dawn of Speech”

Axel G. Ekström¹ 

[1] *Speech, Music & Hearing, KTH Royal Institute of Technology, Stockholm, Sweden.*

Biolinguistics, 2024, Vol. 18, Article e14285, <https://doi.org/10.5964/bioling.14285>

Published (VoR): 2024-04-26

Handling Editor: Bridget Samuels, University of Southern California, Los Angeles, USA

Corresponding Author: Axel G. Ekström, Speech, Music & Hearing, KTH Royal Institute of Technology, Lindstedtsvägen 24, 114 28 Stockholm, Sweden. E-mail: axeleks@kth.se

Abstract

Recent literature argues that a purportedly long-standing theory—so-called “laryngeal descent theory”—in speech evolution has been refuted (Boë et al., 2019, <https://doi.org/10.1126/sciadv.aaw3916>). However, an investigation into the relevant source material reveals that the theory described has never been a prominent line of thinking in speech-centric sciences. The confusion arises from a fundamental misunderstanding: the argument that the descent of the larynx and the accompanying changes in the hominin vocal tract expanded the range of possible speech sounds for human ancestors (a theory that enjoys wide interdisciplinary support) is mistakenly interpreted as a belief that all speech was impossible without such changes—a notion that was never widely endorsed in relevant literature. This work aims not to stir controversy but to highlight important historical context in the study of speech evolution.

Keywords

speech production, evolution of speech, vocal tract, primatology, miscitation

1. Introduction: Wherefrom Theories?

In human speech production, the voice “source” from the vocal folds of the larynx is “filtered” in the supralaryngeal vocal tract (SVT) by the imposition of narrow constrictions using the various articulators, including the jaw, lips, velum, palate, and tongue (Fant, 1960), resulting in variations in the resulting resonance frequencies, termed formants. Because essential features of vocal anatomy are largely preserved across mammals, such fundamentals of speech acoustics have served as starting points for literature on the



evolution of speech capacities (de Boer & Fitch, 2010; Ekström & Edlund, 2023; Fitch et al., 2016; Lieberman et al., 1969, 1972; Negus, 1949).

This text discusses at length claims as to historical sources within the science of speech evolution as presented by Boë et al. (2019) in a recent publication. This work—*Which way to the dawn of speech?* (hereafter the “Dawn of speech” review), published in *Science Advances*—ascribes several views to influential works and researchers, and claims that evidence refutes “laryngeal descent theory”, purportedly an influential theory of speech evolution. The purpose of this text is to evaluate the accuracy of the attributions made there. Three claims are presented and claimed to refute “laryngeal descent theory”, (1) “laryngeal descent is not uniquely human”, (2), “laryngeal descent is not required to produce contrasting formant patterns in vocalizations”, and (3), “living nonhuman primates produce vocalizations with contrasting formant patterns.” Here, I investigate the source material claimed to form the basis and support for the theory. Because the terminology “laryngeal descent theory” does not appear anywhere in the relevant literature prior to the “dawn of speech” review (cf. Barney et al., 2012; Boë et al., 1999, 2002, 2017; Carré et al., 1995; de Boer, 2009, 2010; de Boer & Fitch, 2010; Fitch, 2000, 2010; Fitch et al., 2016; Fitch & Giedd, 1999; Fitch & Reby, 2001; Lieberman, 1984, 1991, 2006, 2007, 2011, 2012, 2017; Lieberman & Crelin, 1971; Lieberman et al., 1969, 1972, 1992; Lieberman & McCarthy, 1999, 2015; Lieberman et al., 2001; McElligott et al., 2006; Negus, 1949; Nishimura, 2005; Owren et al., 1997; Takemoto, 2008), I will refer to “laryngeal descent theory” in quotation marks. The purpose of this text is not to further enflame tensions between researchers in this field, which has historically been unfortunately contentious (see e.g., Lieberman, 2007, 2012). However, it is in the interest of the integrity of our science of language evolution, that errors in attribution be publicly acknowledged, discussed, and corrected.

2. Privileged Sounds of Speech?

Early work on the topic of speech evolution took inspiration from the emerging “quantal theory of speech”. Stevens (1972) originally posited his theory as a principled explanatory model toward explicating non-random patterns of speech sounds observed in the world’s languages. Quantal theory held vowels [a], [i], and [u] were articulated in “stable” regions of phonetic space: they could be straightforwardly produced, even if articulation was imprecise. Decades of research on stop consonants such as [p t k b d g] further indicated that formant transitions may also facilitate the perception of those phonemes (Delattre et al., 1955; Dorman et al., 1977; Kewley-Port, 1982; Liberman et al., 1967). In perception, too, “quantal” vowel phonemes appear privileged. Because of the relative extremity of dispersions of formants for [i] and [u], they have long been believed particularly salient cues for perception (Fitch, 1994; Nearey, 1978). Most recently, research by Friedrichs (2017) has shown that “quantal” vowels so designated by Stevens,

can be more reliably perceived at high fundamental frequencies compared to other vowel phonemes. It was this apparently privileged nature of “quantal” speech sounds, that led to researchers positing a meaningful relationship between those phonemes, and the evolution of human speech capacities.

3. Limits of Monkey Speech

In adult humans, the tongue root is descended into the pharynx, and the tongue, rounded in shape, is positioned in both the pharyngeal and oral cavities. The tongues of nonhuman primates, meanwhile, are flat in shape, and contained almost wholly in the oral cavity (Iwasaki et al., 2019; Negus, 1949; Takemoto, 2008). As a result, while the principal musculature of the primate tongue is well preserved across species, innervation of the musculature results in different vector forces in humans versus non-human primates (de Boer & Fitch, 2010; Takemoto, 2008). The variable most crucial to human speech evolution and phonetic range is the shape and position of the tongue inside the SVT, and the shape of the SVT itself, which throughout the course of human development acquires a distinct right-angle bend at or around its midpoint where the oral and pharyngeal cavities meet (Fitch & Giedd, 1999; Lieberman & McCarthy, 1999; Lieberman et al., 2001; Vorperian et al., 2005, 2009). Lieberman and colleagues (1969, 1972) predicted that non-human primates, reflecting the short and narrow pharynges and flat tongues contained in the oral cavity, could not achieve the midpoint area discontinuities required for the articulatory stability that characterized “quantal” vowels [a], [i], or [u] (Lieberman et al., 1969, 1972). This prediction was most recently supported through work by Fitch et al. (2016), who reported a “vowel space” based on macaque articulatory states that did *not* include these vowel phonemes.

While recent research has documented “[u]-like” calls for a variety of non-human species, including chimpanzees (Grawunder et al., 2022), orangutans (Ekström et al., 2023) and even non-primates such as domestic cats (Schötz, 2020), available evidence suggests these are articulated in manners distinct from that of human speakers, and that they may not meet the criteria of articulatory stability laid out by Stevens (1972, 1989). Observations of “hooting” chimpanzees illustrate that such calls are often produced with visible protrusion and rounding of the lips (Grawunder et al., 2022; Parr et al., 2005) beyond the degree observed in human speakers. Most importantly, however, primate vocal tracts do not allow for human-like production of that vowel. Morphological analyses of the flat tongues of primates like chimpanzees (Takemoto, 2008) and baboons (Boë et al., 2017) reveals that freedom of motion is mainly in protrusion and retraction, precluding many of the rapid movements actively employed in speech. These qualities would thus reflect different SVT configurations compared with human speakers (de Boer & Fitch, 2010; Ekström, *in press*). Grawunder et al. (2023) have recently shown that chimpanzees also produce “[a]-like” calls. Inspection of the animals producing such “bark” calls reveals

articulation with a markedly “flared” oral cavity, i.e., a lowered jaw, beyond the level of human [a]. The long faces of and comparative morphology of the temporomandibular joints (which join the jaw and cranium) allow chimpanzees and other primates to lower the jaw significantly beyond human capacities. These [u]-like and [a]-like calls, thus, resemble their would-be human vowel primarily in acoustic domains, but are not articulated identically.

Perhaps most importantly, nowhere in the literature has an animal been reported as producing a call corresponding to human [i]. This is particularly significant, because [i] has been shown to be an optimal vowel for normalizing vocal tract lengths—believed a central component of how humans can readily “translate” phonetic content between speakers with highly disparate vocal tracts (i.e., children and adults) (Fitch, 1994). Even the most liberal interpretation, thus, indicates that non-human animals possess a limited phonetic range. Even if a non-human primate had a “human brain”, they would not be capable of producing the range, variety and nuances of signals that characterize the speech of all normally developing modern humans. It was for this reason that Lieberman (e.g., 2006, 2007, 2012, 2017) reasoned that a rudimentary system of speech sounds must have evolved in human ancestors, before acting as a novel selection pressure for improved speech communication, toward the modern human SVT.

4. Why the Long Face?

The “roughly equal” proportions of the modern human vocal tract stand in stark contrast to the comparative craniofacial morphology of extant hominids: in an adult chimpanzee, the oral tract is more than twice the length of the height of the pharynx (Nishimura, 2005). The hominin fossil record reveals a long-running pattern of increasingly orthognathic faces, with the anteriormost section of the face being progressively pulled in toward the rest of the skull. Early australopiths such as *Australopithecus anamensis* had the long faces characteristic of modern extant great apes; by the emergence of *H. erectus* some ~2 Mya, however, the craniofacial form of human ancestors was markedly “flattened” in comparison. These changes may reflect dietary shifts away from hard foods (Lieberman, 2011; see also Ledogar et al., 2016).

Like chimpanzees, “long-faced” human ancestors would have possessed a limited range of speech sounds (e.g., Lieberman et al., 1972). Exactly when in the course of human evolution the capacity to produce the full extend of human vowel space emerged has been explored on several occasions (Barney et al., 2012; Lieberman, 2007, 2012; Lieberman et al., 1972; Laitman et al., 1992). Early attempts (Lieberman & Crelin, 1971; Lieberman et al., 1972) were based on the angle of the basicranium—an assumption later shown to be uninformative with the revelation that the tongue roots and larynges of developing humans continue to descend after the point of stabilization of cranial flexure

(Fitch & Giedd, 1999; Lieberman, 2007; Lieberman & McCarthy, 1999). Later attempts were based on a disparate assumption.

Assuming that a roughly equal relationship between horizontal and vertical sections of the SVT is optimal for generating a greatest possible range of speech sounds (Carré et al., 1995, 2017; de Boer, 2010; Stevens, 1972), it is in theory possible to estimate where in the body the larynx would need to be placed to achieve this relationship at resting state conditions. Such estimates of Neanderthal neck lengths (reported in Lieberman & McCarthy, 2015; Lieberman, 2006, 2007, 2012) indicate that assuming roughly 1:1 SVT_H – SVT_V proportions, the Neanderthal larynx would have to be placed inside the thorax, opposite the 3rd thoracic vertebrae. This does not resemble the vocal anatomy of any extant primate. Similar conclusions were reached independently by Barney and colleagues (2012), who also showed a more limited Neanderthal vowel space for the “La Ferrassie 1” Neanderthal. Unlike earlier reconstructions based on the flexure of the basicranium, these more recent reconstructions have yet to be refuted in the relevant literature.

According to these results, Neanderthals would have been precluded from achieving “fully modern” human speech (i.e., be limited to a more restricted range of vowels). That is, according to these estimates Neanderthals (and by extension other non-human primates) would have limited to a less extensive range of speech sounds, including vowels [ɪ], [æ], and [ɛ] (the vowels in “bit”, “cat” and “bed”) but not [a], [i], or [u]. The conclusion that Neanderthal anatomy could not accommodate the morphological prerequisites of the full extent of modern human vowel space (Barney et al., 2012; Lieberman, 2006) would suggest that the modern human SVT—though not necessarily speech or language itself (Dediu & Levinson, 2013; Johansson, 2015; Lieberman, 2007)—emerged after the phylogenetic split with Neanderthals. This point of nuance is particularly pertinent to the framing of the “Dawn of speech” review, which purports to move back the “origins” of speech-centric evolution.

5. Seeking a Smoking Gun: Where Is the Support for “Laryngeal Descent Theory”?

5.1 Narrative Shift: The Sources Tell a Different Story

In the “Dawn of speech” review, two sources are cited for the central tenets of “laryngeal descent theory”: Lieberman et al. (1969) and Lieberman (2007). The first—*Vocal tract limitations on the vowel repertoires of rhesus monkey and other nonhuman primates*—is a seminal empirical work within the discipline of speech evolution. In the article, the phonetic capacities of a rhesus macaque are explored on the basis of acoustic-phonetic computational methods combined with vocal tract length estimates measured from a diseased monkey specimen. Importantly, however, the 1969 article makes note of the fact

that “... nonhuman primates lack a pharyngeal region where the root of the tongue forms a moveable wall.” That is, the relevant work recognizes the importance of the variable shape of the SVT and make no claims as to the “descent of the larynx [being] required for speech”.

The second purported source, *The evolution of human speech: Its anatomical and neural bases* (Lieberman, 2007) is a cohesive review of speech evolution research at the time. However, the 2007 review, like other works by Lieberman (e.g., Lieberman, 1984, 1991, 2012; Lieberman et al., 1972) emphasizes the roughly equal SVT_H – SVT_V , phylogenetic descent of the tongue into the pharynx, and the “quantal” vowels. Summarizing that evolution of both anatomical and neural domains was necessary to explicate the range and nuance of human speech, Lieberman (2007, p. 39) wrote: “The chimpanzee lacks a supralaryngeal vocal tract capable of producing the ‘quantal’ sounds which facilitate both speech production and perception and a brain that can reiterate the phonetic contrasts apparent in its fixed vocalizations.” Sources given in the “dawn of speech” review for claims central of “laryngeal descent theory” claim the opposite.

5.2 Comparative Anatomy Confirms Human Uniqueness

The uniqueness of human vocal morphology was recognized by the surgeon Victor Negus (1949, p. 198), who believed the human vocal tract had acquired its unique characteristics for “... purposes of speech.” Because the reconfiguration renders modern humans more susceptible to choking on food compared with other animals—a seemingly counter-intuitive development antithetical to the primary Darwinian imperative of improving odds of survival—Negus reasoned that this change must have been driven by selection pressures for improved speech communication. This line of reasoning was taken up by Lieberman (1984, 2012) who consistently cited Negus’ work and concurred with his reasoning (e.g., Lieberman et al., 1972). It is noteworthy that Negus, who himself had no training in acoustic phonetics nonetheless reached conclusions that were later supported by independent speech-centric research (Baer et al., 1991; Carré et al., 1995, 2017; Fitch, 2000; Lieberman et al., 1972, 1992; Russell, 1928; Stevens, 1972, 2000).

5.3 Computer Simulations in Search of Efficiency Recreate Human Vocal Tracts

The profound significance of the human reconfigured vocal tract and expanded pharynx with regard to speech production was starkly illustrated through extensive modeling efforts by Carré and colleagues (1995, 2017). The authors found that when computed to identify, step by step, the most efficient configurations for achieving the maximum range of signal variability (i.e., the extent of the resulting vowel space), simulations consistently produce an SVT with independently controllable oral and pharyngeal cavities, roughly equal in length—i.e., one mirroring that unique to modern humans. In nature, only

modern (adult) humans possess SVTs even remotely meeting this criterion. Chimpanzees (Nishimura, 2005) and baboons (Boë et al., 2017) have oral cavities more than twice the length of their pharyngeal cavity, resulting from a combination of a long face and high larynx.

Conclusions by Carré et al. have since been exhaustively replicated (Carré, 2004; Carré et al., 2017), and independently verified through simulations by de Boer (2010, p. 351): "... there is an optimal larynx height at which the largest range of signals can be produced ... at this height, the vertical and horizontal parts are approximately equally long." Contextually, it is worth noting that work by Boë et al. (1999, 2002) has disputed the role of the pharynx in speech evolution. However, the Boë series rests on the basis of a computer model that distorts any input shape into one having the phonetic potential of the adult humans upon which the model was built (de Boer & Fitch, 2010; Lieberman, 2012). Applying the Boë algorithm, SVTs of mice, reptiles, birds, or even inanimate objects would all be shown to possess the human range of phonetic potential. Results of these simulations cannot realistically support any model of speech evolution. Valid computer simulations support the view that the human vocal tract has markedly greater phonetic potential compared to that of non-human mammals, and that this increase in potential involves the ready execution of [a], [i], and [u] (Carré et al., 1995, 2017; de Boer, 2010).

5.4 Languages Differ—but Humans Everywhere Exploit the Same Principles

While the number and features of permissible speech sounds vary greatly across spoken languages (Maddieson, 1984; Moran & McCloy, 2019), the selections of sounds are highly nonrandom (see e.g., Lindblom 1996, p. 66). The exact organizational underpinnings of phonological systems have been the subject of a large number of research works (de Boer, 2010; Diehl, 2008; Lindblom, 1996; Lindblom & Engstrand, 1989; Liljencrants & Lindblom, 1972; Lindblom, & Maddieson, 1988; Stevens, 1972, 1989); while this work has not traditionally been associated with evolution per se (but see Lindblom, 2000), they illustrate the prominence of maximally contrastive vowel sounds. For example, in three-vowel spoken languages, those three vowels tend to be [a], [i], and [u] (the vowels in "ma", "see", and "do") (Moran & McCloy, 2019). That is, the vowels originally designated "stable" by Stevens are the most common vowel sounds across the world's spoken languages. An essentially "human" vocal tract anatomy, thus, was *necessary* for all spoken languages in existence to be produced the way they are, sound the way they do, and be perceived as reliably as they are.

6. A Theory That Never Was

In the “Dawn of speech” review, Boë et al. (2019) summarize that, ‘The global thrust of [Lieberman’s] papers is that laryngeal descent is required for speech’ (Boë et al., 2019, p. 3). This interpretation has been presented previously by Boë et al. (2002, pp. 456–466), who summarized that the conclusion of the Lieberman-Crelin Neanderthal vocal tract dimension estimates had been that Neanderthals “could not speak”. However, the theoretical body of work pioneered by Lieberman and colleagues (1969, 1972; Lieberman & Crelin, 1971) held that unconfigured SVTs (i.e., those of primates, and those inferred for Neanderthals) were limited to “non-quantal” vowels. Lieberman and Crelin did not, however, claim that the lack of a descended larynx would render the speaker “speechless” (Boë et al., 1999). This earlier misinterpretation elucidates the framing of the “dawn of speech” review. Namely, the core of arguments presented therein are based on the interpretation that historically influential literature on the evolution of speech argued that laryngeal descent was the variable crucial to human speech capacities. Rather, authoritative sources consistently emphasized that “a low larynx does not signify an SVT that can produce the full range of human speech” (Lieberman, 2006, p. 278) and that “the key to the evolution of the human SVT involves the descent of the tongue not the larynx, which is carried down into the throat as the tongue moves down into the pharynx and is reshaped” (Lieberman, 2012, p. 612). Considering this review of pertinent historical evidence, let us now turn to the first claim (1) made in the “Dawn of speech” review—the claim that “laryngeal descent is not uniquely human” constitutes an argument against “laryngeal descent theory.”

7. The Descended Larynx and Descending Larynx: Apples and Oranges

Boë et al. (2019) are correct insofar as an undescended larynx allows for the production of a range of speech sounds. Importantly, however, this is not—and never was—in dispute. Indeed, it has long been recognized in relevant literature that nonhuman animal SVTs are underexploited with regards to phonetic range (e.g., de Boer & Fitch, 2010; Fitch, 2000; Fitch et al., 2016; Lieberman et al., 1972, 1992). The claim that “the analysis of primate VT shapes leads Lieberman to relate the human ability to articulate contrasting vowels to the large pharyngeal cavity” (Boë et al., 2019, p. 3) is inconsistent with available sources. As early as 1972, Lieberman and colleagues argued that “... the chimpanzee vocal tract [...] have the anatomic ability that would allow [the animal] to produce a number of vowels that in human speech are “phonemic” elements, i.e., sound contrasts that convey linguistically meaningful information” (Lieberman et al., 1972, p. 299).

The human lowered larynx was never argued to be “required” for all speech production; sources cited for this claim in the “Dawn of speech” review, state the opposite. To

an extent the confusion is understandable. Namely, research over the last two decades has found that a variety of animals readily employ temporary laryngeal descent while vocalizing (Fitch, 2010; Fitch & Reby, 2001; McElligott et al., 2006; Weissengruber et al., 2002). However, while Fitch (2010, p. 317) has argued that "... many of the vocal tract configurations required for speech, are attainable by nonhuman mammals via dynamic vocal tract reconfiguration", this assessment is based on a misleading surface-level similarity and the implication that anatomical evolution was unnecessary is straightforwardly refuted. First, modern humans have no need to dynamically reconfigure their vocal tract in order to achieve the full range of speech sounds; all normally developing humans do so readily, without any need for drastic or effortful changes to anatomy. Second, an animal's reconfiguration of vocal tract anatomy via temporary laryngeal descent does not bestow a human SVT, nor the capacity to rapidly and fluidly apply variable stricture and discontinuity inside the SVT (Carré et al., 1995, 2017; de Boer, 2010; de Boer & Fitch, 2010; Lieberman, 2012; Lieberman et al., 1972, 1992; Stevens, 1972). Thus, the first claim (1) made in the "Dawn of speech" is irrelevant with regards to phonetic capacities: the would-be critique is seemingly based on a conflation of anatomical prerequisites necessary for speech production in a general sense (present across non-human mammals), versus the full range of *human* phonetic capacities in particular. While the notion of articulatory sensitivity is central to the works of Lieberman and colleagues, the "dawn of speech" review does not engage with the concept.

8. Monkey Business: Recent Studies Are Consistent With the Theory That Was

We may now reexamine claims (2) and (3) in the context of two works, which are referred to in the "Dawn of speech" review, as explicit refutations of "laryngeal descent theory".

8.1 What Bored Monkeys (Do not) Say About Speech Origins: On Fitch et al. (2016)

The "Dawn of speech" review claims that work by Fitch et al. (2016)—which investigates the range of SVT configurations ("vowel space") available to a macaque monkey—is a refutation of "laryngeal descent theory." While the macaque was shown to be in theory capable of a range of human vowel phonemes—through vocal tract configurations employed in vocalizing, lip smacking, yawning (Everett, 2017)—that range did not extend to the "quantal" vowels designated as uniquely robust (Ekström, *in press*; Lieberman, 2017). The work illustrates that the macaque, even incorporating yawning gestures (phonemically unrealistic SVT configurations, unlikely to be incorporated into real-life speech due to demands on muscular effort and time to execution) does not achieve the

SVT configurations corresponding to [a], [i] or [u]. The work by Fitch et al. (2016) is consistent with the account of speech evolution that holds that the reconfiguration of the hominin vocal tract was necessary to produce the full range of speech sounds.

8.2 Brainy Baboons: On Boë et al. (2017)

The “dawn of speech” review argues that the observation that a baboon (*Papio .spp*) “vocalic proto-system” is evidence against “laryngeal descent theory” (Boë et al., 2017, 2019). This conclusion, too, is based on the historically inconsistent assumption that there was ever widely held support for the idea that the human permanently descended larynx was “required for speech”. Importantly, too, while the work on baboon capacities purportedly shows that baboons can achieve [u]-like calls in the same manner as humans (i.e., using the tongue and pharynx), there are several problems with this assertion. Namely, there is no positive evidence that baboons produce such vowel-like quality using identical lingual gestures to humans; the authors conjecture this connection to anatomy based on audio data. The authors’ dissection of baboon vocal tracts confirms lingual degrees of freedom mainly in protrusion and retraction. Like all nonhuman primates, baboons have high larynges and short-and-narrow pharynges and flat tongues contained in the oral cavity. The anatomical data refutes the authors’ conjecture of baboons employing identical articulatory gestures to those observed in modern humans: in baboons, such articulation is not possible. Stricture resulting in an [u]-like formant dispersion is thus more likely to result from a significant retraction of the tongue body, causing the oropharynx to narrow: This is markedly less efficient than “true” [u], produced by all normally developing humans. More poignantly, however, because the sources cited do *not* claim that primates are “incapable of producing contrasting sounds”, the work does not constitute evidence against the body of theoretical works outlined above. Rather, this work, too, is consistent with the account of speech evolution that holds that the reconfiguration of the hominin vocal tract was necessary to produce the full range of speech sounds.

9. Wrong Way to the “Dawn of Speech”

The “dawn of speech” review states that three claims constitute evidence against “laryngeal descent theory”. Against the background discussed above, we find that the first—(1) “laryngeal descent is not uniquely human”—is irrelevant, as the mechanisms of dynamic larynx lowering are phylogenetically and morphologically distinct from the permanent laryngeal descent in human ontogeny. Dynamic lowering does not bestow the animals with human vocal tracts; and (1) appears to be based on the incorrect summary that “laryngeal descent was [believed to be] required for speech” in speech-centric evolutionary sciences (cf. Lieberman et al., 1972), which is not consistent with historical sources. The

second and third claims—(2) “laryngeal descent is not required to produce contrasting formant patterns in vocalizations”, and (3) “living nonhuman primates produce vocalizations with contrasting formant patterns”—are unquestionably true. However, the “dawn of speech” review paradoxically interprets decades of work that recognized the truth of these claims as arguing *against* them.

The misinterpretation can be traced back to a fundamental misconception: namely, the claim that laryngeal descent and concomitant reconfiguration of the hominin vocal tract expanded the range of speech sounds available to human ancestors (for which there is universal support across phonetic sciences) is taken to mean that all speech was impossible without such reconfiguration. The first remains well supported by decades of relevant research in speech production, comparative vocal anatomy, speech acoustics, speech perception and phonology. However, there is little evidence of widely held support for the second—now, or at any other point in the history of our discipline. As such, the “Dawn of speech” review constitutes an unfortunate distortion of the relevant scientific literature. It is vital to the integrity of our science that such errors in attribution can be publicly acknowledged, discussed, and corrected.

Funding: The results of this work will be made more widely accessible through the national infrastructure Språkbanken Tal under funding from the Swedish Research Council (2017-00626).

Acknowledgments: The authors have no additional (i.e., non-financial) support to report.

Competing Interests: The author has declared that no competing interests exist.

References

- Baer, T., Gore, J. C., Gracco, L. C., & Nye, P. W. (1991). Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels. *The Journal of the Acoustical Society of America*, *90*(2), 799–828. <https://doi.org/10.1121/1.401949>
- 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 of London. Series B, Biological Sciences*, *367*(1585), 88–102. <https://doi.org/10.1098/rstb.2011.0259>
- 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., Becker, Y., Rey, A., & Fagot, J. (2017). Evidence of a vocalic proto-system in the baboon (*Papio papio*) suggests pre-

- hominin speech precursors. *PLoS One*, 12(1), Article e0169321.
<https://doi.org/10.1371/journal.pone.0169321>
- Boë, L. J., Sawallis, T. R., Fagot, J., Badin, P., Barbier, G., Captier, G., Ménard, L., Heim, J.-L., & Schwartz, J. L. (2019). Which way to the dawn of speech?: Reanalyzing half a century of debates and data in light of speech science. *Science Advances*, 5(12), Article eaaw3916.
<https://doi.org/10.1126/sciadv.aaw3916>
- Carré, R., Lindblom, B., & MacNeilage, P. F. (1995). Rôle de l'acoustique dans l'évolution du conduit vocal humain. *Comptes Rendus de l'Académie des Sciences. Série II. Fascicule b, Mécanique*, 320, 471–476.
- Carré, R. (2004). From an acoustic tube to speech production. *Speech Communication*, 42(2), 227–240. <https://doi.org/10.1016/j.specom.2003.12.001>
- Carré, R., Divenyi, P., & Mrayati, M. (2017). *Speech: A dynamic process*. De Gruyter.
- de Boer, B. (2009). Acoustic analysis of primate air sacs and their effect on vocalization. *The Journal of the Acoustical Society of America*, 126(6), 3329–3343. <https://doi.org/10.1121/1.3257544>
- de Boer, B. (2010). Modelling vocal anatomy's significant effect on speech. *Journal of Evolutionary Psychology*, 8(4), 351–366. <https://doi.org/10.1556/JEP.8.2010.4.1>
- de Boer, B., & Fitch, W. T. (2010). Computer models of vocal tract evolution: An overview and critique. *Adaptive Behavior*, 18(1), 36–47. <https://doi.org/10.1177/1059712309350972>
- Dediu, D., & Levinson, S. C. (2013). On the antiquity of language: The reinterpretation of Neandertal linguistic capacities and its consequences. *Frontiers in Psychology*, 4, Article 397. <https://doi.org/10.3389/fpsyg.2013.00397>
- 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>
- Diehl, R. L. (2008). Acoustic and auditory phonetics: The adaptive design of speech sound systems. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 363(1493), 965–978. <https://doi.org/10.1098/rstb.2007.2153>
- 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>
- Ekström, A. G., & Edlund, J. (2023). Evolution of the human tongue and emergence of speech biomechanics. *Frontiers in Psychology*, 14, Article 1150778.
<https://doi.org/10.3389/fpsyg.2023.1150778>
- 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 20th International Congress of Phonetic Sciences* (pp. 3076–3080). Guarant International.
<https://doi.org/10.31234/osf.io/8aeh4>
- Ekström, A. G. (in press). *Phonetic potential of primate vocal tracts and the legacy of Philip Lieberman (1934–2022)*. *American Journal of Primatology*. <https://doi.org/10.31219/osf.io/hd48k>

- Everett, C. (2017). *Yawning at the dawn of speech: A closer look at monkey formant space*. Retrieved November 6, 2023, from http://calebeverett.org/uploads/4/2/6/5/4265482/commentary_on_fitch_et_al.pdf
- Fant, G. (1960). *The acoustic theory of speech production*. Mouton.
- Fitch III, W. T. S. (1994). *Vocal tract length perception and the evolution of language* [Doctoral thesis]. Brown University.
- Fitch, W. T. (2010). *The evolution of language*. Cambridge University Press.
- 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., & Reby, D. (2001). The descended larynx is not uniquely human. *Proceedings of the Royal Society of London: Series B, Biological Sciences*, 268(1477), 1669–1675. <https://doi.org/10.1098/rspb.2001.1704>
- Fitch, W. T., De Boer, B., Mathur, N., & Ghazanfar, A. A. (2016). Monkey vocal tracts are speech-ready. *Science Advances*, 2(12), Article e1600723. <https://doi.org/10.1126/sciadv.1600723>
- Friedrichs, D. (2017). *Beyond formants: Vowel perception at high fundamental frequencies* [Doctoral thesis]. University of Zurich.
- 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), Article 20200455. <https://doi.org/10.1098/rstb.2020.0455>
- Grawunder, S., Uomini, N., Samuni, L., Bortolato, T., Girard-Buttoz, C., Wittig, R. M., & Crockford, C. (2023). Correction: ‘Chimpanzee vowel-like sounds and voice quality suggest formant space expansion through the hominoid lineage’(2021), by Grawunder et al. *Philosophical Transactions of the Royal Society B*, 378(1890), Article 20230319. <https://doi.org/10.1098/rstb.2023.0319>
- 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>
- Johansson, S. (2015). Language abilities in Neanderthals. *Annual Review of Linguistics*, 1(1), 311–332. <https://doi.org/10.1146/annurev-linguist-030514-124945>
- 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>
- Laitman, J. T., Reidenberg, J. S., & Gannon, P. J. (1992). Fossil skulls and hominid vocal tracts: New approaches to charting the evolution of human speech. In J. Wind, B. Chiarelli, B. Bichakjian, A. Nocentini, & A. Jonker (Eds.), *Language origin: A multidisciplinary approach* (pp. 385–397)

- Ledogar, J. A., Dechow, P. C., Wang, Q., Gharpure, P. H., Gordon, A. D., Baab, K. L., Smith, A. L., Weber, G. W., Grosse, I. R., Ross, C. F., Richmond, B. G., Wright, B. W., Byron, C., Wroe, S., & Strait, D. S. (2016). Human feeding biomechanics: Performance, variation, and functional constraints. *PeerJ*, 4, Article e2242. <https://doi.org/10.7717/peerj.2242>
- 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, 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., McCarthy, R. C., Hiimae, 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. (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. *Linguistic Inquiry*, 2(2), 203–222.
- Lieberman, P., Crelin, E. S., & Klatt, D. H. (1972). Phonetic ability and related anatomy of the newborn and adult human, Neanderthal man, and the chimpanzee. *American Anthropologist*, 74(3), 287–307. <https://doi.org/10.1525/aa.1972.74.3.02a00020>
- Lieberman, P. (1984). *The biology and evolution of language*. Harvard University Press.
- Lieberman, P. (1991). *Uniquely human: The evolution of speech, thought, and selfless behavior*. 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. (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. <https://doi.org/10.1016/j.wocn.2012.04.001>
- Lieberman, P., & McCarthy, R. C. (2015). The evolution of speech and language. In W. Henke & I. Tattersall (Eds.), *The role of speech in language* (pp. 83–106). Springer.
- Lieberman, P. (2017). Comment on “Monkey vocal tracts are speech-ready”. *Science Advances*, 3(7), Article 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*, 48(4), 839–862. <https://doi.org/10.2307/411991>

- Lindblom, B., & Maddieson, I. (1988). Phonetic universals in consonant systems. In L. Hyman & C. Li (Eds.), *Language, speech and mind* (pp. 62–78). Routledge.
- Lindblom, B., & Engstrand, O. (1989). In what sense is speech quantal? *Journal of Phonetics*, *17*(1-2), 107–121. [https://doi.org/10.1016/S0095-4470\(19\)31516-5](https://doi.org/10.1016/S0095-4470(19)31516-5)
- Lindblom, B. (1996). Role of articulation in speech perception: Clues from production. *The Journal of the Acoustical Society of America*, *99*(3), 1683–1692. <https://doi.org/10.1121/1.414691>
- Lindblom, B. (2000). Developmental origins of adult phonology: The interplay between phonetic emergents and the evolutionary adaptations of sound patterns. *Phonetica*, *57*(2-4), 297–314. <https://doi.org/10.1159/000028482>
- Maddieson, I. (1984). *Patterns of sounds*. Cambridge University Press.
- McElligott, A. G., Birrer, M., & Vannoni, E. (2006). Retraction of the mobile descended larynx during groaning enables fallow bucks (*Dama dama*) to lower their formant frequencies. *Journal of Zoology*, *270*(2), 340–345. <https://doi.org/10.1111/j.1469-7998.2006.00144.x>
- Moran, S., & McCloy, D. (2019). *PHOIBLE 2.0*. Max Planck Institute for the Science of Human History. <https://phoible.org>
- Nearey, T. (1978). *Phonetic features for vowels*. Indiana University Linguistics Club.
- Negus, V. E. (1949). *Comparative anatomy and physiology of the larynx*. Heinemann.
- Nishimura, T. (2005). Developmental changes in the shape of the supralaryngeal vocal tract in chimpanzees. *American Journal of Physical Anthropology*, *126*(2), 193–204. <https://doi.org/10.1002/ajpa.20112>
- Owren, M. J., Seyfarth, R. M., & Cheney, D. L. (1997). The acoustic features of vowel-like grunt calls in chacma baboons (*Papio cyncephalus ursinus*): Implications for production processes and functions. *The Journal of the Acoustical Society of America*, *101*(5), 2951–2963. <https://doi.org/10.1121/1.418523>
- Parr, L. A., Cohen, M., & Waal, F. D. (2005). Influence of social context on the use of blended and graded facial displays in chimpanzees. *International Journal of Primatology*, *26*, 73–103. <https://doi.org/10.1007/s10764-005-0724-z>
- Russell, G. O. (1928). *The vowel: Its physiological mechanism as shown by the X-ray*. Ohio State University Press.
- 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* (pp. 319–347). Springer.
- Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David, Jr & P. B. Denes (Eds.), *Human communication: A unified view* (pp. 51–66). McGrawHill.
- Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, *17*(1-2), 3–45. [https://doi.org/10.1016/S0095-4470\(19\)31520-7](https://doi.org/10.1016/S0095-4470(19)31520-7)
- Stevens, K. N. (2000). *Acoustic phonetics*. MIT Press.

- 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>
- 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., Ziegert, A. J., & 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>
- 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>