SOCIAL SCIENCES

Which way to the dawn of speech?: Reanalyzing half a century of debates and data in light of speech science

Louis-Jean Boë[1]*, Thomas R. Sawallis[2]*, Joël Fagot[3,4], Pierre Badin[1], Guillaume Barbier[1,5], Guillaume Captier[6], Lucie Ménard[7,8], Jean-Louis Heim[9,10†], Jean-Luc Schwartz[1]

Recent articles on primate articulatory abilities are revolutionary regarding speech emergence, a crucial aspect of language evolution, by revealing a human-like system of proto-vowels in nonhuman primates and implicitly throughout our hominid ancestry. This article presents both a schematic history and the state of the art in primate vocalization research and its importance for speech emergence. Recent speech research advances allow more incisive comparison of phylogeny and ontogeny and also an illuminating reinterpretation of vintage primate vocalization data. This review produces three major findings. First, even among primates, laryngeal descent is not uniquely human. Second, laryngeal descent is not required to produce contrasting formant patterns in vocalizations. Third, living nonhuman primates produce vocalizations with contrasting formant patterns. Thus, evidence now overwhelmingly refutes the long-standing laryngeal descent theory, which pushes back "the dawn of speech" beyond ~200 ka ago to over ~20 Ma ago, a difference of two orders of magnitude.

INTRODUCTION

Full language is unique to and universal in humans, where it is universally transmitted by vocal speech. Animals do communicate in various ways, including with vocal calls, but the structural complexity, flexibility, and integration of speech and language in humans are vastly greater than anything found in other species. Understanding the gulf between the human and animal systems and specifically how the modern human system, language, emerged evolutionarily through extinct hominins from the lesser systems of our hominid ancestors has been called the hardest problem in science (*1*). Our aim in this article is to present both a schematic history and the state of the art in research on vocal communication of nonhuman primates to better illuminate the emergence of human speech—a limited but crucial component of the broader topic of language emergence—by using a comparison of phylogeny and ontogeny and to add a contribution that has only recently become possible.

The study of speech evolution is necessarily multidisciplinary, involving at a minimum paleoanthropology, primatology, and speech science, itself already a conglomeration of phonetics, anatomy, acoustics, human development, and more. The long process of interweaving these typically independent disciplines over such a complex research problem to build a consensus will inevitably entail controversy as well as progress, but analyzing scientific controversies before they are resolved allows us to observe "science in action" (*2*): how practicing researchers collect data, develop hypotheses, advance interpretations, and commit to them before theories reach a general consensus. The nature and history of that interweaving is the foundation of this article. We hope that readers will find it instructive to reflect on the paths and processes involved.

Specifically, in the "The development of research in speech evolution and the LDT" section, we present the beginnings of speech evolution research with the laryngeal descent theory (LDT), which very strongly affected the nature of research on this topic with its claim that only modern humans could produce fully contrasting vowel qualities. In the "LD is not uniquely human: Ethology, primatology, bioacoustics, and primate and animal communication" section, we discuss how ethology and primatology addressed primate communication within the LDT framework yet progressively accumulated evidence incompatible with LDT. In the "Vowel qualities can contrast without a low larynx: Fundamentals of speech acoustics" section, we explain the major elements of articulatory and acoustic analysis of speech production and show how they strongly imply that laryngeal descent (LD) is not required to produce contrasting vowels. In the "Uniform tubes cannot explain primate vocalizations: The acoustic capacities of nonhuman primates" section, we present a large amount of data on nonhuman primate vocalizations, showing that they do involve contrasting vowel qualities. Last, in the "Lessons learned from the LD controversy" section, we analyze the lessons learned and some new overall perspectives on the problem of speech emergence.

More generally, our intent is to fully examine the current state of knowledge about the evolutionary roots of human vowel production. Here, we will not enter into the vigorous and important debates on other aspects of language evolution, including syntax, lexicon, gesture, and neurolinguistics, nor do we address other crucial aspects of speech, such as consonants, phonological representations, syllabic organization, speech perception, or neuromuscular control. At least two reasons justify our tight focus on vowel production: First, vowels are the core of speech production and are required to effectively transmit consonantal acoustics, which together enable a phonologically encoded lexicon, which is then subject to syntax. Thus, vowel production enjoys logical primacy, because no other aspect of spoken language has any utility until the articulatory ability to produce vowels is established. Second, LDT's claim that only fully modern humans can produce contrasting vowels has been argued as restricting all aspects of language emergence to the past 200,000 years, thus bolstering claims of the flowering of full human language within the past 100,000 to 70,000 years (*3*). If LDT is

[1]Université Grenoble Alpes, CNRS, Grenoble INP, Institute of Engineering Univ. Grenoble Alpes, GIPSA-lab, Grenoble, France. [2]New College, The University of Alabama, Tuscaloosa, AL, USA. [3]Brain and Language Research Institute, Aix-Marseille University, Aix-en-Provence, France. [4]Cognitive Psychology Laboratory, Centre National de la Recherche Scientifique and Aix-Marseille University, Marseille, France. [5]School of Speech Pathology and Audiology, Université de Montréal, Montréal, Québec, Canada. [6]Anatomy Laboratory, Montpellier University, Montpellier, France. [7]Laboratoire de Phonétique, Université du Québec à Montréal, Montréal, Québec, Canada. [8]Center for Research on Brain, Language, and Music, Montréal, Québec, Canada. [9]Muséum National d'Histoire Naturelle, Paris, France. [10]Institut de Paléontologie Humaine, Paris, France.
*Corresponding author. Email: louisjean.boe@orange.fr (L.-J.B.); tom.sawallis@gmail.com (T.R.S.)
†Deceased.

refuted, and contrasting vowel qualities were available earlier, that opens for reexamination the time frames for emergence of all subsequent aspects of language. Thus, in our view, these questions of vowel production are not trivial, but foundational, for the field of language evolution.

THE DEVELOPMENT OF RESEARCH IN SPEECH EVOLUTION AND THE LDT

Primate vocalizations have been understood for over a century as a crucial element in the study of language emergence, and the LDT played a foundational role in the organization and development of this field in the past 50 years.

Why link speech emergence and primate vocalizations?

Comparative method is a standard tool in evolution research. Cognition regarding the more abstract aspects of language has been studied by comparisons with, for instance, birds, cetaceans, and dogs, as well as nonhuman primates. However, the need for anatomical similarity means that investigation of speech emergence can best be restricted to our closer primate kin, the hominoids (apes) and cercopithecoids (Old World monkeys).

As a basis of comparison, and given the suggestion that evolution changed the hominin line faster than other primates (*4*), we can reasonably posit that the communicative behaviors of the other living primates may be relics of behaviors shared by our last common ancestors (*5*). Because communicative behavior does not fossilize, we can metaphorically take their communication as fossils of prior communicative abilities (*6*) and compare the vocal articulation, the acoustic results, and the anatomical structures that enable them.

Neural anatomy and function, both central and peripheral, differ between modern humans and other living primates, but neurology is subordinate to physics in determining the sound production capacity of a species' anatomy, so anatomy and its acoustic effects are our focus. Here, the fundamental difference is that our larynx is lower in the throat (relative to the cervical vertebrae), opening up a large pharyngeal cavity. For a half century, this difference was theorized to preclude the production of contrasting vowel qualities by any but anatomically modern *Homo sapiens* (AMHS) and thereby to restrict the emergence of language until after their appearance, some ~200 thousand years (ka) ago. Over the decades, objections and even controversy arose regarding that dominant theory. We here aim to explain the steps whereby the growing evidence from speech science techniques, by many other researchers as well as our own team, slowly broke through that alleged limit, showing the existence of proto-vowels and systemic precursors of human speech in other living primates, thereby pushing back "the dawn of speech" beyond ~200 ka ago to more than ~20 million years (Ma) ago.

Pioneering work of Lieberman

Philip Lieberman, then at the Haskins Laboratories, was among the first researchers to use the new techniques and concepts of the modern era of speech science to study nonhuman vocalizations and, more generally, to address the topic of speech emergence, and he did so in a truly pioneering fashion. He also did so with auspicious timing; by the 1950s, two scientifically documented projects to teach speech and language to a chimpanzee had failed [for a review, see (*7*)].

Methodological advances

From 1968 to 1971, Lieberman and his colleagues published several groundbreaking articles using methodologies that, with technological updates, are still standard. To begin, Lieberman (*8*) recorded various chimpanzee, gorilla, and rhesus macaque calls; analyzed their acoustics with spectrograms (the standard tool for visualizing acoustic speech analysis since the late 1940s); and then used the acoustic results to make inferences about their vocal tract (VT) anatomies and their anatomical differences. A few years later, he and his colleagues used the same methodology to investigate human infant cries from birth to 4 days (*9*).

Next, Lieberman and colleagues (*10*) determined the shape of the cavity in a nonhuman primate VT, at that time using plaster casts and x-rays on a rhesus macaque. They used "a computer-implemented model of the supralaryngeal vocal tract" to explore the degrees of freedom in the monkey's articulation. This model allowed them to explore the animal's acoustic potential by simulating variations in its VT shape, for comparison with the articulation and acoustics of humans.

Last, with Crelin (*11*), he used casts of fossil skulls to infer Neanderthal VT anatomy and cavity shapes. These were extensively compared to both adult and infant humans (Edmund Crelin was a specialist on neonatal anatomy) and then used to estimate the Neanderthal acoustic potential as had just been done with the rhesus macaque. Using Peterson and Barney's classic survey of human vowels (*12*), this potential was then compared to human productions.

It bears noting that in these articles, Lieberman and his colleagues covered the requisite foci for investigating human evolution: fossil hominids, human growth and development, and living nonhuman primates. However, the key innovation is that through both colleagues and a framework that was multidisciplinary, he applied to these investigations classic speech science techniques that remain fundamental today: acoustic signal analysis, anatomical description, and articulatory and acoustic modeling.

Together, these studies represent an extremely powerful research paradigm, drawn directly from the core understanding of speech science, that articulatory information is transmitted in the acoustic speech signal. Specifically, (i) from the recorded calls of live animals, one can make anatomical inferences; (ii) from anatomical data on casts of extant species, fossils, or cadavers, one can make acoustic inferences; and (iii) by appropriate comparisons of anatomy and acoustics, one can make extrapolations regarding both the ontogeny and the phylogeny of speech.

Conclusions drawn

The conclusions drawn by Lieberman from these studies, taken together, are as follows.

First

> We can … infer that the energy concentrations in the spectrogram of gorilla Kathy's vocalization reflect the transfer function of her supralaryngeal vocal tract in the schwa configuration. (*8*)

> Our data indicate … that the nonhuman primates would not be capable of producing human speech even if they had the requisite mental ability. Unlike man, the nonhuman primates do *not* appear to change the shape of their supralaryngeal vocal tracts by moving their tongues during the production of a cry. (*8*)

From analysis of acoustic calls and simulations drawn from VT cadavers, Lieberman finds that outside modern human adults, evidence is negligible for volitional deviation from a VT configured as a uniform tube. A uniform tube has the same cross-sectional area from one end to the other (i.e., from glottis to lips in a VT) and is typically modeled as a cylinder. When adult humans vocalize through a VT

configured as an approximately uniform tube, the result is the schwa, /ə/, but in running speech, humans characteristically use precise tongue, jaw, and lip gestures to achieve nonuniform VT configurations. These allow production of all the vowels dispersed through human articulatory space and documented since the 1880s in the International Phonetic Alphabet (IPA) (*13*). The claim by Lieberman is thus that only humans have the capacity to modify their VT shape and the corresponding acoustic resonances (*14*), while all other primates would produce only a schwa-like vocalization in the center of the vowel space with little to no capacity for such modification.
Second,

> [T]he rhesus monkey is inherently incapable of producing the range of human speech. … The nonhuman primates lack a pharyngeal region like man's, where the cross-sectional area continually changes during speech. The inability of apes to mimic human speech (Kellogg, 1968) is thus an inherent limitation of their vocal mechanisms. (*10*)

The analysis of primate VT shapes leads Lieberman to relate the human ability to articulate contrasting vowels to the large pharyngeal cavity, which is the main anatomical distinction between human adults and the others investigated. Crucially, the phylogeny of AMHS included LD, where the laryngeal cartilages and the glottis separated from the hyoid bone and moved lower (relative to the cervical vertebrae), thus creating a substantial vertical cavity where there had been very little. Tongue movements were capable of changing the cross-sectional area of the pharynx and also affecting the cross section of the preexisting oral cavity. This form and relation between oral and pharyngeal cavities were alleged to distinguish humans from other primates by enabling the full acoustic space, and specifically /i a u/, the most extreme vowels in the space, for the first time. These vowels (also termed point vowels) mark the boundaries of the vowel systems in all human languages, as noted in the IPA system and verified in the UPSID (UCLA Phonological Segment Inventory Database) database of several hundred representative human languages (*15*). Thus, the high larynx and resulting small pharynx of other primates as compared to humans would prevent them from producing these crucial articulatory and acoustic landmarks, which Lieberman finds to be absent from the restricted vowel space of the rhesus macaque.
Third,

> Newborn human infants, like nonhuman primates, do not execute any maneuvers of their supralaryngeal vocal tracts during vocalizations except for gross laryngeal maneuvers. The shape of their supralaryngeal vocal tract appears to approximate a uniform cross-section, schwalike, configuration. (*9*)

> The newborn infant, like a nonhuman primate, thus lacks a pharyngeal region that can vary its cross-sectional area. … [T]he newborn infant, like the nonhuman primates, is restricted by the limitations of his vocal apparatus. (*9*)

Lieberman and his coauthors note that recordings of newborns resemble those of nonhuman primates, in that they present the schwa-like pattern of formants (amplified regions in the spectra) typical of a uniform tube articulation. From an anatomical point of view, AMHS infants also share the high larynx and small pharynx of living nonhuman primates, with a larynx in the standard mammalian position (*16*), implying that they, like those primates, should be unable to articulate the requisite /i a u/ for speech. They would attain this capacity over the course of development, but only in late childhood, after a notable lowering of the larynx. Later, following anatomical work by others (*17*, *18*), but without detailing clear acoustic justification, Lieberman further specified (*19*–*21*) that the horizontal oral and vertical pharyngeal cavities of the supralaryngeal VT (often termed SVTh and SVTv) must be proportioned about equally, a 1:1 SVTh/SVTv ratio, for fully human speech.
Fourth,

> We have previously determined by means of acoustic analysis that Newborn humans, like nonhuman primates, lack the anatomical mechanism that is necessary to produce articulate speech. … We can now demonstrate that the skeletal features of Neanderthal man show that his supralaryngeal vocal apparatus was similar to that of a Newborn human. … It appears that the ontological development of the vocal apparatus in Man is a recapitulation of his evolutionary phylogeny. (*11*)

Lieberman and Crelin reconstructed the Neanderthal VT and determined that the larynx was still high. This gave Neanderthals a VT structure similar to that of newborns and would leave them equally unable to articulate the requisite /i a u/ for speech. This was confirmed through articulatory modeling and acoustical analysis, as had been done for rhesus macaque. Moreover, this led them to find human phylogeny to be reflected in AMHS ontogeny and further confirmed that only AMHS adults were fully capable of speech.
Fifth,

> The data suggest that speech cannot be viewed as an overlaid function that makes use of a vocal tract that has evolved solely for respiratory and deglutitious purposes; the skeletal evidence of human evolution shows a series of changes from the primate vocal tract that may have been, in part, for the purpose of generating speech. (*8*)

> The human speech-output mechanism thus should be viewed as part of man's species-specific language endowment. (*10*)

The global thrust of these papers is that LD is required for speech, but how does that fit with human evolution more broadly? As a skeptic of speech "as an overlaid function," Lieberman seems in 1968 to oppose explaining speech by what later became known as "exaptation" (*22*), the use of an anatomical structure for a function other than that for which it was developed by natural selection. That is, because speech could not be accomplished with the structures meant for respiration and nourishment, speech had to be recent because it had to wait for LD to occur in AMHS. Then, in 1969, he adds that speech is required for language; thus, without the universal point vowels /i a u/ to mark the full human vowel space, there could not be full human speech, and without speech, there could be no full human language. In particular, language could not have emerged earlier than speech nor speech earlier than LD, so emergence of both speech and language must have occurred recently, after LD as a component of the emergence of AMHS, which has generally been thought to date to about 200 ka ago.

These appear to us as the core of Lieberman's claims: that human infants, nonhuman primates, and pre-AMHS hominids can only

produce schwa-like vowels, while AMHS adults alone can articulate the full array of vowels because of the 1:1 ratio of SVTh and SVTv resulting from their descended larynx and large pharynx, and that the key steps in phylogeny occurred with LD first, speech emergence next, and then language emergence.

LDT: Dissemination and influence

With a protocol solidly based on acoustic and anatomical observations, with marked advances through pioneering articulatory-acoustic modeling and with strong claims in the important topic of speech evolution, Lieberman's conclusions became a touchstone, accepted as essentially fact, and came to be commonly referred to as laryngeal descent theory (LDT). It was taught in textbooks and routinely disseminated in publications on the origin of language and seemed so evident that it was taken as canonical, even axiomatic. It was taken by scientists as explaining the failure to teach speech and language to home-raised chimpanzees. As documented below, researchers tended to overlook criticism of his conclusions, even serious arguments regarding articulatory capabilities of Neanderthals (*23*–*25*) or infants (*26*, *27*), because these arguments could be discounted as tangential to the core of the theory, the vocalizations of nonhuman primates. Moreover, although LDT generally lacked support from speech researchers (see below), the LDT became an early foundational tenet of a complex school of thought claiming a recent, sudden, and simultaneous appearance of speech and language in AMHS (*3*). In our view (*28*, *29*), LDT is a paradigmatic example of a "contagious idea" understandable through the "epidemiology of representations" framework developed in anthropology and cognitive sciences by Sperber (*30*, *31*) to explain how some theories spread more quickly and broadly than others, sometimes before the formation of any consensus on their scientific validity:

> A certain number of researchers were convinced [of LDT] by the simplicity of the reasoning, its coherence, and its explanatory power, showing the parallel of ontogeny and phylogeny while contrasting [modern] humans both with Neanderthals and apes, and this in spite of its [LDT's] high implausibility: women have a distinctly higher larynx than men, but they easily contrast [i a u]. (*28*)

Years of silence

A decade elapsed after the foundational articles in the LDT (1968–1971) before further "speech-oriented" studies appeared regarding nonhuman primate vocalizations for several reasons. First, the theory itself explicitly denied the possibility of spectral differentiation of vocalizations outside modern humans, which naturally demotivated researchers from doing analyses to find any in other primates. The LDT also provided support for those advancing gesture, not speech, as the origin of language (*32*–*34*).

Next, such studies present well-known practical problems: Acoustical analysis of recordings made in the field is quite challenging because of the weak signal-to-noise ratio, and laryngeal function is less stable in monkeys and apes than in human speech and so presents a noisier, harsher structure and makes formant measurement difficult (*6*, *35*, *36*).

Last, the technology available for the analyses was underpowered. Manual production of analog spectrograms with visual determination of formants and fundamental frequencies (i.e., the vocal fold vibration rate) was both time consuming and imprecise. Linear predictive coding (LPC) analysis (*37*), the modern alternative to traditional sound spectrography (see the "Vowel qualities can contrast without a low larynx: Fundamentals of speech acoustics" section), became common in speech analysis in the late 1970s, but it took another 10 years before it was adopted in ethology, bioacoustics, primatology, and animal communication. Moreover, the central notion of formants, advanced not only by Lieberman (*8*) but also by Andrew (*38*) and Richman (*39*) for the study of nonhuman primates, was "virtually ignored in work with nonhuman species" (*40*). It was 1988 before formant detection with LPC was used for characterizing vervet alarm calls (*40*) and then later for gorilla double grunts (*41*) and rhesus macaque coos and screams (*42*). LPC finally became more common when used as of the late 1990s to extract formant values while investigating the correlation between VT length (VTL) and body size in macaques and baboons (*43*–*45*).

One can imagine that progress might have been faster if the evolution of speech examined within the framework of animal communication had involved more phoneticians, but ironically, the speech science community mainly stayed silent on the topic. The principal exception was Ohala (*46*, *47*), whose insightful but controversial work from an ethological viewpoint, presented below, attracted little response.

LD IS NOT UNIQUELY HUMAN: ETHOLOGY, PRIMATOLOGY, BIOACOUSTICS, AND PRIMATE AND ANIMAL COMMUNICATION

Researchers beyond the speech community continued to work on animal communication as related to primate vocalization while accepting (at least provisionally) the LDT and concentrating their work along several major themes, which bear discussion.

Research in ethology

Animal communication is research worthy on its own merits, and bioacousticians were making important strides there through the work of their leading scholar, Marler [e.g., see (*48*)] and others. These touched on the use of vocal and other communication for classic ethological concerns such as mating, food, aggression, territorial defense, and social rank. Because Lieberman's work had apparently closed the door to research on human-like speech in living species, primatologists with interests in vocal communication fell back on studies with traditional ethological themes, sometimes incorporating theoretical concepts and technological processes from speech science according to their interests, abilities, and needs.

Marler (*49*) suggested that animal cries were more than expression of emotion and that they had symbolic content. In the wild, various monkey species produce different calls in response to different situations or objects. These calls have been termed functionally semantic or functionally referential because the calls elicit the same behavioral response from hearers as they would have given had they personally experienced the same stimulus as the callers. For instance, alarm calls that differentiate between terrestrial and aerial predators elicit the appropriate differing defensive responses, even from those who cannot see the predator, as shown by playback experiments. The details of these studies have been extensively reviewed, but their implications are still under debate [for a recent review, see (*50*)].

Vervets have provided the classic example of alarm call differentiation since Struhsaker (*51*) first documented that different predators elicited different alarm calls. A variety of observational and experimental studies subsequently investigated whether those calls "refer" to predators similarly to the way human vocabulary involves semantic reference to objects and situations (*40*, *52*–*54*). The apparent referential quality leads Riede and Zuberbühler (*45*) to suggest that "formant modulation is the result of active vocal filtering used by the monkeys to encode semantic information."

The evidence for acoustic-semantic relationships in Diana monkey and vervet alarm calls and vocalizations was based initially and partly on perceptual evidence of playback experiments. As for acoustic analysis of vocalizations, one of the first, on vervet calls (*55*), showed the methodological difficulties of detecting formants. Owren and Bernacki (*40*) characterized these calls as "relatively inaccessible to spectrographic examination due to their noisy broadband structure" and therefore used the more powerful LPC analysis for their spectral measures of vervet alarm calls, apparently the first use of LPC on primate vocalizations.

None of these studies directly refer to LDT, but we will see below that a number of them pointedly remark that some of the vocalizations studied present clear acoustic evidence of differences in VT form that would permit the production of "phonetic contrasts."

Multiparametric discrimination: Characterizing the repertoire of vocalizations

We have noted a number of primate communication studies, similar among themselves in selecting a set of around 10 time- or frequency-domain parameters as entry points for acoustic analysis. These acoustic data then serve as input for a discriminant function analysis, with results generally projected on a two-dimensional (2D) plot with the two leading discriminant function analysis axes. The acoustic parameters used as foundation often involve duration, mean fundamental frequency (F0), frequency range, mean deviation of the peak frequency, distribution of frequency amplitude of quartiles, and presence or absence of noise. This experimental design, with similar acoustic parameters, has also been used to study human speech, for instance, with the goal of detecting emotions in vocal productions [e.g., (*56*, *57*)]. In general, the approach is well adapted to categorizing signals that can be sorted into a limited number of predefined sets, as is the case sorting primate vocalizations (e.g., grunts, barks, and screams for baboons) according to ethologically distinct situations of behavior and communication (*58*, *59*).

Such a design was used by Fischer and colleagues (*60*) on the barks of female chacma baboons to look for variation correlating with context, predator type, and individuality. In Guinea baboons, Maciej and colleagues (*61*) found six call types, two distinct categories of screams and two of grunts, as well as barks and wahoos. In their study of black howler monkeys (*Alouatta pigra*), Briseño-Jaramillo *et al.* (*62*) found that the two-component space of their discriminant analysis showed clusters corresponding to nine call types comparable to those found in prior studies. Benítez *et al.* (*63*) analyzed the wahoos of male geladas (*Theropithecus gelada*) during ritual chases with rival males. Their discriminant analysis showed the wahoos can be used by other geladas to assess the qualities of a potential rival or a potential mate.

These studies are all implicitly external to the question of LD. While they are entirely based on acoustic measurements, including spectral data, they neither explicitly address the filter function of the VT nor refer to formants and nor consider whether differences found in the acoustic spectrum might stem from different VT configurations. We can nonetheless infer that this is the case, because the authors successfully categorize primate vocalizations using acoustic parameters including the spectrum.

A new proposal: LD for body size signaling

In 1984, Ohala (*46*), a leading scholar in phonology and phonetics with interests in ethological aspects of communication, hypothesized that a set of "disparate phenomena" from speech, language, and human and animal biology is related through an underlying "frequency code." He notes a variety of effects that can signal a human individual's potential dominance or cooperativity in an interaction, including sexual dimorphism of VT anatomy, which is essentially absent until puberty. Thereafter, an extra bout of LD lengthens the male's VT another 15 to 20%, resulting in lower resonance frequencies. He takes this as evidence against three proposed explanations for LD (adaptation to upright bipedalism, adaptation to reduced prognathism—i.e., shorter "snout"—and, pertinent here, adaptation for speech and language) for three reasons: First, females have no speech deficit associated with their less descended larynx. Second, the males' final LD coincides not with any increased need for speech but with the start of competition for mates, as is the case for sexual dimorphism in most other species. Third, many species without upright posture, with snouts, and without language or speech have similar sexual dimorphism of the VT, including gorillas, howler monkeys, elephant seals, various birds, and, most spectacularly, the bird of paradise, which has ~80 cm of trachea coiled in its ~25-cm body. For Ohala, this is evidence pointing to LD not as an adaptation for speech as claimed by LDT but as one manifestation of a cross-species recognition that lower frequencies signal larger vocalizers and higher frequencies signal smaller vocalizers—the "frequency code" as size signaler. This signal is then indirectly associated with various essentially ethological meanings involving threat, social rank, and interaction intent, each meriting appropriate investigation.

In 1997, Fitch (*43*) confirmed part of Ohala's hypothesis, showing that formant frequencies allow estimation of both VTL and body size in rhesus macaques from 1 to 9 years old. He disagreed with Ohala, though, that the evidence refuted Lieberman's claims:

> Although Ohala initially offered this proposal as a refutation of Lieberman's "phonetic expansion" hypothesis [Ohala, 1984], the two are compatible, with size exaggeration providing a preadaptation for the evolution of speech. (*64*)

Meanwhile, although researchers continued to consider that nonhuman primates were incapable of producing differentiated vowel qualities, they began in the early 2000s to present evidence that LD was not "uniquely human" and thus to challenge the idea that it was both necessary and sufficient for the emergence of speech. Permanently descended larynges were documented first in deer (*65*) and soon thereafter in Mongolian gazelles (*66*), in a variety of felid species (*67*), and eventually in chimpanzees (*68*) and other primates (*69*). Berthommier *et al.* (*6*) proposed a global portrait of comparative data on larynx height and larynx descent in human and nonhuman primates. Fitch (*70*) discovered and documented a more widespread process of dynamic, temporary lowering of larynx, hyoid, and tongue root in a similarly diverse set of mammals: goats, dogs, pigs, and cotton-top tamarins.

Of course, any putative need of LD for speech and then language does not explain its presence in these diverse mammal species. The most important hypothesis in that regard holds that body size exaggeration provides the selective impetus for LD. Fitch's expansion (*64*) of Ohala's original suggestion (*46*) is paraphrased by Rendall *et al.* as:

> Fitch argues that a descended larynx lengthens the vocal tract, thereby lowering the formant frequencies and signaling larger body size, with attendant advantages in social competition. … In short, humans' descended larynx reflects a history of sustained selection for reliable body-size cuing, and its descended position was only secondarily co-opted for a language function. (*71*)

Fitch eventually concludes that dynamic LD is characteristic of all mammalia and generalizes this ability explicitly to our hominid ancestry, to chimpanzees (our closest living relative), and, implicitly, much further. He also makes explicit the pertinent inference regarding prehuman VTs and abilities:

> [E]ven the earliest vocalizing hominids could attain a vocal tract configuration adequate for producing many clear, comprehensible phonemes by simply doing what all mammals do: reconfiguring the vocal anatomy while vocalizing. (*72*)

Body size exaggeration would thus have selected for LD, whether permanent or dynamic, in many mammals, including hominids, where it was then available for exaptation for speech. Researchers working on this idea are not studying speech phenomena per se but "vocal features evolutionarily linked to expression of body size and sex (fundamental and formant frequencies)" (*5*), and this new paradigm is epistemologically distinct. In this paradigm, VTL is the key to animal communication as an individualized indicator of size and weight (*64*, *71*), whether honest or exaggerated (*73*, *74*).

This proposal rests on a foundation of hypotheses, demonstrations, and corroborations regarding the central role played by formants in animal communication. In contrast with previous studies where parallels with linguistics are evident, these studies present similarities with forensic speech analysis, such as the estimation of talker body size and height from recordings. While these concerns are not central to speech communication per se, they contribute to progress in studies on speech evolution by bringing the attention of animal communication researchers to the acoustics of formants, a key parameter in speech. The important findings obtained with this new paradigm are as follows:

1) Animal communication researchers recognized that the spectral maxima in primate vocalizations are simply formants. This affirms the formant detections of Lieberman (*8*, *10*) and subsequent early studies (*38*, *39*) and is further buttressed by later formant measurements made to estimate VTL. The spectral patterns of baboon grunts are sufficiently similar to vowels in speech that Owren *et al.* (*44*) and Rendall (*75*) term them vowel-like, even without the stable voicing and fundamental frequency that are characteristic of full vowels.

2) Primates perceive differences between signals with contrasting formant structures (*52*, *76*–*80*), and replay studies using resynthesized calls with controlled formant variations show that they precisely monitor formant modifications (*81*).

3) VTL and body size (height and/or weight) are related, but the relation is not straightforward. They correlate in macaques from 1 to 9 years (*43*), in humans from 2 to 25 years (*17*), and in mammals generally (*82*). However, Rendall *et al.* (*71*) later tested the relation between vocalizations and body size using human vowels and "the vowel-like grunts of baboons, whose phylogenetic proximity to humans and similar vocal production biology and voice acoustic patterns recommend them for such comparative research" in an approach using "body size and voice-acoustic allometry." Notable among Rendall *et al.*'s conclusions: that there is a "mismatch between F0 and body size in both species" (*71*) and that "[i]n humans, formant variation is correlated significantly with speaker height but only in males and not in females" (*71*). Last, Hatano *et al.* (*83*) measured VTL directly in magnetic resonance images (MRIs) of adults and found that while there were weak correlations between VTL and formant frequencies, neither reveals body size, as VTL does not correlate significantly with body size once full adult stature is attained, so "the vocal tract length does not reflect the body height."

In brief, while research continues, the evidence currently seems to support Ohala's contention that LD in humans is simply one instance of size signaling for ethological reasons. As such, it occurs in many species and is not, as LDT claims, a uniquely human adaptation for speech.

Vowel quality differences interpreted as articulatory maneuvers

Over the course of the years, studies sporadically produced evidence that was recognized as inexplicable under LDT's claim that primate VTs were restricted to schwa-like configurations. Early on, Richman (*39*) had presented spectrograms showing that geladas could produce a wide variety of vocalic and consonantal contrasts similar to those of speech. However, the lack of anatomical details about the recorded monkeys and of any statistical data about the acoustics made the results seem anecdotal or prospective.

In 1984, Seyfarth and Cheney (*55*) noted that vervets produce spectrally differentiated grunts in four different social contexts and potentially for different social purposes. Then, in 1988, Owren and Bernacki (*40*) gave evidence of "phonetic contrasts" between vervet snake and eagle alarm calls. They attribute these contrasts to VT shape modification, contrary to Lieberman's predictions (*84*) that contrasts would be found, but due to laryngeal source characteristics or nonoral cavities (i.e., air sacs). Owren and Bernacki conclude that

> Our findings, together with Seyfarth and Cheney's [1984] data, raise the possibility that vervets may also routinely manipulate vocal tract resonance characteristics during call production. (*40*)

In 1993, Hauser *et al.* (*85*) analyzed audio and video recordings of rhesus monkeys and found evidence of articulatory movements of the jaw and of both opening and protrusion of the lips, all causing formant variation indicating nonuniform VT configurations well beyond what had been predicted by LDT. Shortly thereafter, Owren *et al.* (*44*) found that, while their study found little evidence of active formant manipulation, "the adult female baboon vocal tract is not entirely uniform over its length and can therefore not be exactly matched by an idealized straight-tube resonator."

Together, these studies tend to show that both apes and monkeys modify their VT configurations via articulatory movements that tend to dissociate the spectral patterns (formants) from the characteristics of the laryngeal source (*86*). Still, an important threshold for theory is crossed when Riede and Zuberbühler (*45*) suggest that Diana monkeys use formant transitions in alarm calls to encode semantic information, as humans do in diphthongs, glides, and transitions between vowels and consonants. They are undoubtedly studying the vocalization of nonhuman primates as though it were speech, as is emphasized by their comparison of the alarm call's formants to those of human vowels analyzed by Lee *et al.* (*87*) in adult males and in 10 to 12 year olds, whose VTL is comparable to those of Diana monkeys. From radioimaging and dissections, Riede *et al.* (*88*) then propose a simplified (three-tube + lips) production model and various articulatory movements, including of the jaw, to simulate the formant transitions observed in Diana monkeys' leopard alarm calls. Contrary to LDT predictions, they conclude:

> Diana monkey leopard alarm calls … overlap substantially with the /a/ vowel and the /o/ vowel F1/F2-range of a 10 to 12 years old child with a similar vocal tract length (*88*).

Larynx in Neanderthals

Lieberman's reconstruction of the Neanderthal VT was challenged by numerous researchers. The criticisms centered on the position of the hyoid bone, which was very high relative to the jaw, and would have left it difficult, if not impossible, to either lower the jaw:

> In this position digastrics can only pull jaw up and back into depth of glenoid fossa; force-couple action on jaw is impossible! (*24*)

or to swallow:

> The reconstructed hyoid bone has been placed in a position unlike that occupied by hyoid bones of newborn humans, adult humans, stillborn chimpanzees or adult chimpanzees. In any laryngeal reconstruction, the function of swallowing must be taken into account. The ability of the reconstructed Neanderthal to swallow is discussed in light of a comparative analysis of swallowing in man and the chimpanzee. It is concluded that the statement that Neanderthal was less than fully articulate remains unsubstantiated because it rests on a questionable reconstruction of the larynx. (*23*)

This topic was addressed in detail by Boë *et al.* (*89*). Their study capitalized on a previous publication by Honda and Tiede (*90*) analyzing biometric orofacial data from MRIs of modern human subjects. Honda and Tiede showed a statistical relationship between three basic descriptors of the orofacial anatomy: (i) the palatal distance between two reference points on the palate (the anterior and posterior nasal spines), which also provided a reference "palatal line"; (ii) the oral cavity height, defined as the distance between the lowest point of the mandible (gnathion) and the palatal line; and (iii) the larynx height (LH) defined as the distance between larynx (arytenoid apex) and the palatal line. Thus, for modern humans, the larynx position can be predicted from the other reference points.

While the arytenoid is not preserved in fossils, the other reference points are hence the proposal by Honda and Tiede (*90*) that the LH could be estimated from the visible nasal spines and gnathion on skulls of archaic humans such as Neanderthals. In line with this proposal, Boë *et al.* (*89*) first tested and validated the correlations provided by Honda and Tiede with larger databases (measurements from midsagittal x-rays of Egyptian and South American mummies, courtesy of the Musée de l'Homme, Paris). Then, they applied the method to predict LH for the skulls of two male adult Neanderthals, La Chapelle-aux-Saints and La Ferrassie 1 (*91*), dated in the range of 45,000 to 70,000 years. Both the cranium base and the mandible were preserved in these two fossils. From the LH estimation, Boë and colleagues computed the LH index (LHI) defined as the ratio of LH and palatal distance (similar, though not identical, to the SVTh/SVTv ratio previously discussed). They were able to show that the LHI values for the two Neanderthal skulls were within the range of variation of LHI values for AMHS subjects. Because of prognathism (a more protruding "snout"), the LHI value is less than 1 in Neanderthals and similar to the ratio in 10-year-old AMHS children. Comparable results were obtained with a similar technique by Barney *et al.* (*92*).

These converging studies suggest that larynx position was essentially identical in Neanderthals and AMHS and thus that there is no anatomical basis for suggesting that LD was needed to provide AMHS with phonetic abilities unavailable to Neanderthals. Because of their extinction, of course, it is hard to imagine direct proof that Neanderthals used the point vowels /i a u/, but the evidence here raises serious doubts about Lieberman and Crelin's argument (*11*) that Neanderthals lacked the requisite "anatomical bases" for speech.

To sum up the "LD is not uniquely human: Ethology, primatology, bioacoustics, and primate and animal communication" section, over the course of about 35 years, sufficient evidence accumulated to finally lead primatologists to explicitly question or even abandon LDT and to start to map out the implications of that shift for speech evolution. Anthropologists' opinions also evolved concerning larynx position in Neanderthals. Next, we discuss some of the methodologies that speech science brings to that discussion.

VOWEL QUALITIES CAN CONTRAST WITHOUT A LOW LARYNX: FUNDAMENTALS OF SPEECH ACOUSTICS

In a recent handbook on animal bioacoustics, Fitch and Suthers (*93*) discuss the difficulties biologists encounter trying to adapt the principles and methods of speech research to study animal communication. As an advanced introduction to the topic and to subsequent discussion, we aim in this section to present certain fundamentals of speech production, VT modeling from birth to adulthood, and vowel system organization that we have found crucial for the analysis of primate vocalizations. Many of these ideas are detailed further in classic textbooks [e.g., (*94*, *95*)].

Let us note that the material we present consists also of concepts and methods developed over more than two decades of work by multiple overlapping international multidisciplinary teams (including researchers in phonetics and vowel universals, VT modeling, acoustic speech processing, anatomy, genetics, VT ontogeny, speech development, paleo- and physical anthropology, primatology, and cognition), linking a core group at GIPSA-lab in Grenoble, France, with researchers from many different laboratories. This multidisciplinary collaboration was necessary to reopen the doors to lines of inquiry and research on the emergence of speech that had been effectively barred by the consensus around LDT.

Source-filter theory

The acoustic spectrum and formant structure of the speech signal were made visible by the spectrograph (*96*) and then readable by the acoustic theory of speech production (*97*), the two combining to reveal the relationship between formants and certain key aspects of the VT configurations. This section explains the core of that theory and its instantiation in modern articulatory-acoustic research on speech.

Principles

Fant's acoustic theory of speech production (*97*) is also known as the source-filter theory because it explains speech sounds as arising from glottal vibration as a source signal, which is then modified by the VT as a filter (with the simplifying assumption that minor interactions between the glottal source and the VT are discounted). The speech signal in typical vowels (i.e., voiced oral vowels) comes from the expiratory airflow that initiates or maintains periodic modulation by the aerodynamic effects of its passage through the glottis (the gap between the vocal folds). That source signal is a plane wave that moves from the glottis, through the VT, and out at the lips. It is spectrally simple, with the energy decreasing rapidly up the F0's harmonics (successive multiples of F0, the vocal folds' frequency of vibration). The VT transforms the acoustic spectrum supplied by the laryngeal source and redistributes its

spectral energy by filtering it according to the resonance characteristics implied by the VT's configuration.

Acoustic theory allows determination of the VT's filter characteristics through several analytical steps. First, because the VT is a bent nonuniform tube, the VTL must be evaluated in 2D along the VT's median line in the midsagittal plane, from the glottis to the lips. MRI [e.g. (*98*)] has replaced lateral radiography and tomography [e.g., (*97*)] for this step. Then, the transverse cross-sectional area, the third dimension of the VT, is sampled in planes perpendicular to and along that median line. These sampled areas are reconstituted as "stacked" cylinders aligned axially on their centers, resulting in a straight tube with variable sections [the acoustic effects of straightening are negligible; (*99*)], which is acoustically equivalent to the VT. Plotting those areas gives the area function, specifying the cross-sectional area of each cylindrical element as a function of its distance from the glottis.

From that area function—which is an acoustically sufficient 3D representation of the VT—Fant's theory enables calculation of the acoustic transfer function, which quantifies the filter that the VT applies to the glottal source. The amplified regions in the resulting spectrum are termed formants, labeled Fn and numbered from low to high frequencies (F1 the lowest, F2 next, etc.). They are the key to contrasting vowel qualities in human speech (specifically, the first three formants, F1-F3, characterize the contrasting vowels of human speech, acoustically for researchers and perceptually for listeners). Formants can also be calculated indirectly from a recorded speech signal, either by visual analysis of a classic spectrogram or through digital processing using fast Fourier transform or, as we discuss below, LPC analysis. The analytical keys to the source-filter theory are illustrated in Fig. 1 (presented right to left, in keeping with the century-plus convention in phonetics of presenting the sagittal section from its left, a convention we follow throughout this paper).

Source-filter theory has been widely adopted for research on mammal vocalizations and, more generally, for bioacoustics (*40*, *93*, *100*, *101*). Because the spectral energy in human speech rolls off sharply as frequency rises (as noted above), detection of F4 is uncertain and that of higher formants is rare. In other primates, though, researchers typically detect F6 (*43*) and higher, sometimes up to F12 (*102*). This is possible because the glottal source signal in primate vocalization is regularly very rich up to the high spectral frequencies.

Formant extraction by LPC

LPC was developed in the late 1960s [see (*37*) for historical perspectives] and became common in speech research during the 1970s to allow estimation of the VT transfer function directly from the speech signal. While the mathematical details and LPC's specific strengths, difficulties, and limits are beyond our scope here [see (*103*, *104*), or (*105*) for different treatments of those aspects], LPC exploits the redundancy in the signal by using a sequence of samples of the digitized signal to predict the subsequent samples. This method has the advantage of separately calculating the rapid changes attributable to the laryngeal source from the slower changes in resonance patterns (i.e., formants) due to movement of the VT. This makes it essentially parallel to the source-filter theory (*97*) in separating the effects of the glottal source from the VT

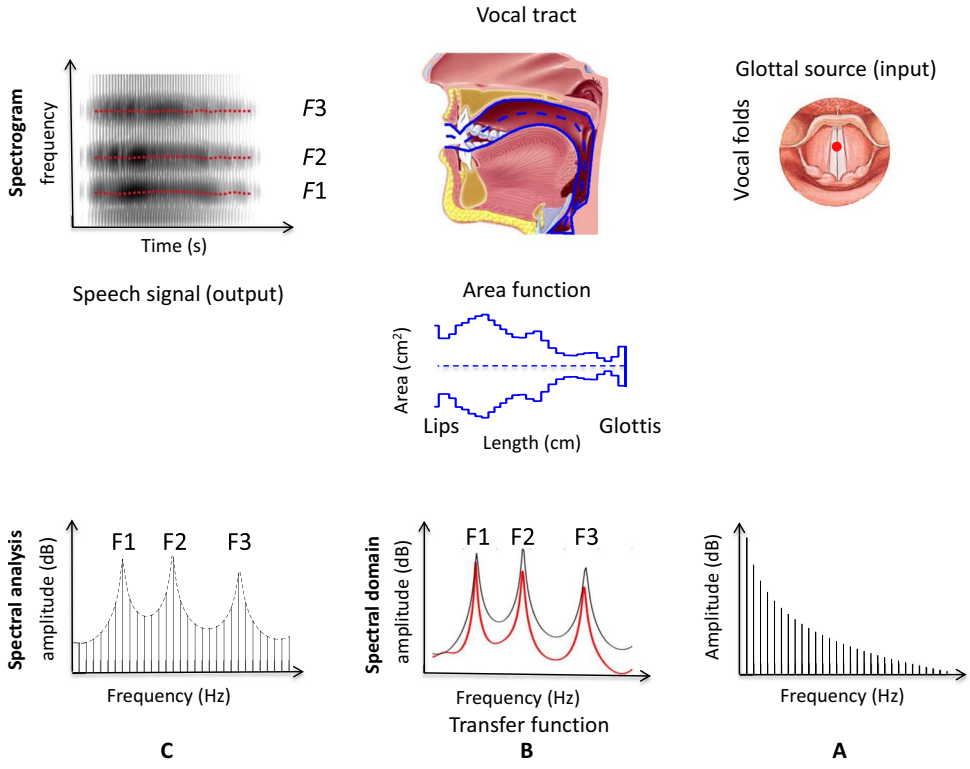

Fig. 1. Source-filter theory. Interpreting vowel spectra as the result of the transformation, by the VT, of the glottal source . (**A**) Source. Top: view from above on the vocal folds; bottom: spectrum of the glottal source signal, with high amplitude in low frequencies. (**B**) Filter. Top: sagittal section of the VT, with the median line (dotted) where VTL is calculated; middle: VT's area function; bottom: VT's acoustic transfer function; black: calculated from the area function; red: extracted by LPC analysis from speech signal. (**C**) Radiated sound. Top: classic spectrogram of a synthesized vowel, showing formants and their peak frequencies over time [here calculated by Praat (*171*)]; bottom: vowel's acoustic spectrum, whose amplitude envelope (dotted line) is imposed on the source signal by the VT transfer function.

as a resonating filter and gives it its power as a tool to isolate and specify formant values. Note, however, that to use LPC successfully, the user must specify the number of formants expected in the signal. Years of experience with LPC have taught researchers to manage these settings for the different talkers and various vowels of human speech, but the settings are difficult to determine for primate vocalizations, and we are just beginning the long effort to master them (*35*). It is perhaps for this reason that, as we will see later, animal communication research did not incorporate LPC analysis until the late 1980s, about two decades after its appearance in speech research.

Relating VT shapes to acoustic resonances

Tubes, cavities, constrictions, and acoustic resonances

VT shapes are spectrally characterized by their acoustic resonances (formants) so that every variation in VT configuration produces variations in the spectrum and in the resulting formant pattern. Contrary to Lieberman's reasoning in his founding papers, the set of formants achievable by a given acoustic tube does not depend directly on the anatomical structures defining its shape (i.e., the oral and pharyngeal cavities) but mainly on its length and the arrangement of acoustic "cavities" and "constrictions" along that length, as will be discussed now.

As a principle, speech science has known at least since Fant (*97*) that production of vowels requires control of the area of the lip opening (Al) and of both the location (Xc) and area (Ac) of a VT constriction—that is, a place in the VT where the narrowing by the tongue toward the palate, velum, or pharyngeal wall defines relatively decoupled acoustic cavities. In the phonetic tradition, the two parameters Xc and Ac define the height and place of articulation of the vowel, and the parameter Al is a correlate of rounding or labialization. Examining these aspects of the three extreme vowels, the observed shapes are /i/ with a large back cavity, a long narrow constriction by the tongue dorsum in the front of the palate, and a small flared opening at the lips (Fig. 2A, top); /a/ with a short constriction low in the pharynx followed by a flared bell shape (Fig. 2B, top); and /u/ with two large closed cavities of roughly equal size, the back one closed by the tongue dorsum around the midpalate and the front one closed at the lips by protrusion with very tight rounding (Fig. 2C, top). Note also that the sensitivity of the three parameters differs across vowels: For /i/, Al can vary as long as the lips are somewhat open, whereas Xc, the constriction location, must be very precise; for /a/, Al is similarly variable, but the ratio of Al/Ac is important for F1 (*106*); and for /u/, both the lip opening and the constriction must be very small, while the constriction location can vary through the middle of the VT (*107*, *108*). In LDT's founding publications, Lieberman's VT simulations (*9*, *10*) apparently did not attain the small, accurate constrictions necessary in the high sensitivity areas of /i u/ or the Al/Ac ratio for /a/.

The classic four-tube model of Fant (*97*) used the component tubes to represent not the anatomical cavities per se (pharynx and oral cavity) but the acoustic cavities characterized by the tongue constriction and the lip opening. In his representation, the tubes representing the lip opening and the constriction are of fixed length, with variable areas representing Al and Ac. The constriction scrolls through the model as a whole, so the model does allow both a constriction at the lips and an internal constriction whose placement is flexible and can be directed from /a/, through /u/, to /i/. It thus instantiates the control parameters of Fant's source-filter theory, as Fant demonstrates with spectra and F1-F5 frequencies for a range of (Xc, Ac, Al) values.

Crucially, the acoustic cavities do not correspond directly to the oral and pharyngeal anatomical cavities, nor is there any need or use for a 1:1 SVTh/SVTv relationship. Instead, the acoustic cavities are defined by jaw opening, tongue shape, and the lip pavilion shape. This erroneous presupposition of equivalence between anatomic (oral and pharyngeal) and acoustic (front and back) cavities is the core flaw of LDT and can be traced to its founding publications.

Maximal acoustic space

Our team contributed an important methodological advance by formalizing the idea of the "maximal acoustic space" (MAS) of a given tube model. The MAS consists of the exhaustive exploration of the range of (F1, F2)—and possibly (F1, F2, F3)—values attainable by the model. For this purpose, we used an n-tube model (*29*) incorporating, but more general than, the four-tube model introduced by Fant (*97*) and determined the MAS in the following way. We set a fixed overall model length l (e.g., the 17.5 cm typical of the VT of an adult male human) and a plausible Ac range for vowel production (i.e., a minimum large enough to avoid consonantal turbulence effects and a maximum small enough to be articulatorily realistic). We divided the tube into the required n component tubes, set areas for each component tube, and calculated the acoustic transfer function to extract F1 and F2 values. Each such model will have $n - 1$ degrees of freedom for the locations of the tube's divisions and n degrees of freedom for the areas the tube, for a total of $2n - 1$ degrees of freedom. Applying a Monte Carlo method, we randomly generated a large number of values for these variables (typically 500,000 sets, for high precision along borders) and thus determined the full extent of the F1-F2 space available to the model specified, i.e., a model of that length with that number of component tubes.

With this technique, we showed that all n-tube models when $n \geq 4$ cover the same area in the F1-F2 space and hence correspond to the same MAS. We also showed that the set of possible (F1, F2) values corresponds to the classical vowel triangle with the uniform tube at the center and /i a u/ at the corners. This shows that it suffices to be able to produce a constriction anywhere inside the tube plus one at the labial end to ensure that the whole set of (F1, F2) values is reachable inside the vowel triangle. This confirms that anatomical details per se—such as the length of the pharyngeal region and the resulting SVTh/SVTv ratio—do not restrict the articulation of any vocalizations inside the vowel triangle: For a given VTL, it suffices to exhaustively vary the three parameters (Xc, Ac, and Al).

These results corroborate Fant's choice of the four-tube model. Our equivalent model (with $n = 4$) gives schematized area functions that illuminate the relationships between the cavities and the formants (see Fig. 3). VT configurations characteristic of the three corners of the MAS show that there is no configuration capable of making these point vowels other than those mentioned previously: for /a/, a discretized horn shape with tube of increasing areas; for /i/, a small open front cavity and a large closed back cavity configured as a Helmholtz resonator (see Fig. 3 for explanation), where the palatal constriction serves as its neck; and for /u/, two closed cavities as Helmholtz resonators with the necks (the constrictions) enclosing approximately equivalent volumes by positioning the tongue near the uvula and closing the lips. The uniform tube corresponds to the schwa, /ə/, around the center of the F1-F2 vowel triangle. Crucially, all vowels of the world's languages can be displayed and positioned precisely inside the MAS, as shown in Fig. 3.

The finding that the MAS forms a triangle with /i a u/ at the extrema constitutes evidence that it is not, as LDT claimed, a particular anatomical configuration (specifically, a lowered larynx enlarging the pharynx or a 1:1 SVTh/SVTv ratio) that permits articulation of these three

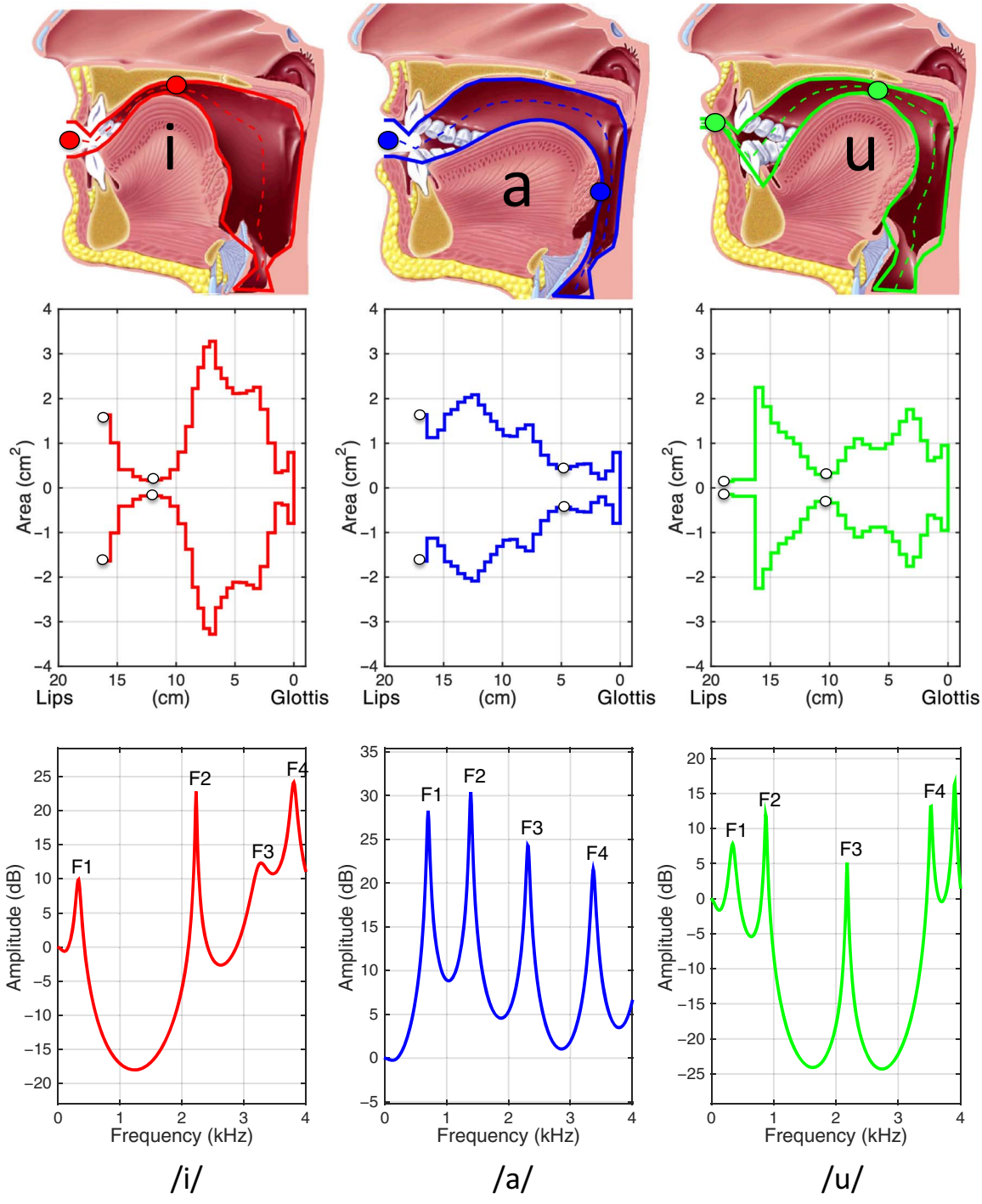

Fig. 2. Production of the three extreme vowels /i/, /a/, and /u/. (**Top**) Sagittal section (with the median line along which the VTL is measured from glottis to lips) with dots indicating the lips (location of Al) and the vowel's constriction (with location Xc and area Ac) along the VT. (**Middle**) Area function, with dots in corresponding locations. (**Bottom**) Acoustic transfer function, with the lowest formant peaks marked (F1 to F4).

vowels, which are nearly universal in human languages. Rather, the intrinsic properties of tube acoustics are exploited via talkers' articulatory control to generate the acoustic triangle and thus furnish the maximum potential for vowel differentiation. That is, knowing only a VT's length (and nothing about its anatomy), one can calculate its MAS, locate a given production therein, and determine its phonetic value and its symbol in the IPA. With appropriately sampled vocalizations (e.g., from a given species), such analysis can show whether the vocalizations cover the MAS and exploit its full potential for contrast. The MAS can thus serve (as we will see below for nonhuman primates) to compare vocalic qualities produced by VTs of differing lengths. For instance, using the data from Peterson and Barney (*12*) (data publicly available in Praat),

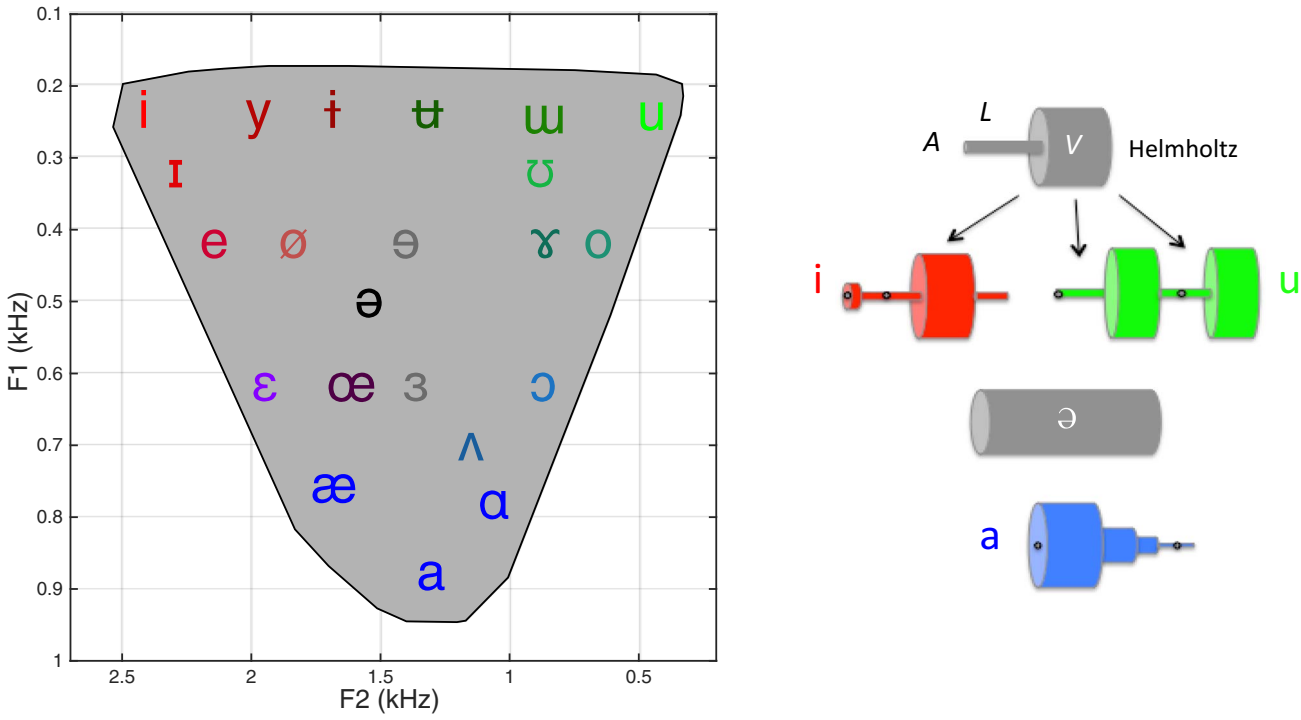

Fig. 3. Four-tube MAS, main IPA vowels, and schematic /i a u/ configurations. The MAS was set at l = 17.5 cm. In the vowel configuration schemas, the dots show the key points of the VT constriction (for Xc and Ac) and the lip opening (Al). A Helmholtz resonator consists of a body of volume V that is extended by a neck of length L and of area A. Because the frequency of a Helmholtz resonator is proportional to $\sqrt{A/(LV)}$, the smaller and longer the neck, and/or the larger the volume, the lower the resonance. The single Helmholtz resonator of /i/ gives it its low F1, and the pair of Helmholtz resonators in /u/ make both F1 and F2 low. Note that the orientation of the F1 and F2 axes in the MAS is standard in speech research to match the preexisting conventional orientation of the vowel triangle in the IPA, defined by tongue position of a speaker facing left. Note also the color scheme of the IPA vowels, which is used here and below for convenience.

we observe that the dispersion ellipses for American English vowels uttered by men, women, and children are well dispersed and cover their representative MASs reasonably well for VTLs of 17, 15, and 12.5 cm, respectively (Fig. 4).

Despite its simplicity (for instance, its lack of distinction of SVTh and SVTv), the MAS generates the entire F1-F2 vowel space possible for a VT of a given fixed length. However, because it also generates forms impossible for the tongue to articulate, it does not allow "inversion" from acoustic (F1, F2) values to retrieve exclusively realistic articulations, as when analyzing recorded vocalizations. To do so, we introduce in the next section a different kind of model incorporating anthropomorphic constraints.

Anthropomorphic articulatory modeling

Expanding on the work of Lindblom and Sundberg (*109*) and Harshman *et al.* (*110*), Maeda (*111*) improved the articulatory modeling of the VT using a principal components analysis of the variability of a number (typically 30) of sagittal sections (Fig. 5A) typically obtained by cineradiography (*112*) of a native talker pronouncing a representative French corpus. (Note that French includes the three extreme vowels, /i a u/, that are reference points for the IPA.) His analysis reduced the partially correlated sagittal view contour points to a set of linearly uncorrelated control parameters closely related to known articulatory variables (e.g., jaw, lips, tongue, and larynx), which then drive the model as command parameters controlling, e.g., vertical larynx position, tongue position, jaw position, and lip opening. For Maeda's model, the seven control parameters account for 70 to 90% of the observed variation in vowel productions (Fig. 5B) (*111*, *113*).

This model naturally suggests a mapping between the model's control parameters and the in vivo articulators, which was later investigated for tongue muscle activity through electromyography by Maeda and Honda (*114*). Confirmation comes from further studies both by electromyography (*115*, *116*) and biomechanics (*117*), which also showed that there existed a simple mapping between activity of the tongue muscles (hyoglossus, anterior and posterior genioglossus, and styloglossus), the tongue parameters of the model, and the F1-F2 pattern.

The anthropomorphism of Maeda's model allowed it to generate the F1-F2 maximum vowel space (MVS) (*118*, *119*). The MVS's range is explored by applying the Monte Carlo method to the seven control parameters (similar to our development of the MAS, discussed above) and then synthesizing and plotting the resulting vowels. The evident similarity of the MVS and the MAS (see Fig. 6 and "Tubes, cavities, constrictions, and acoustic resonances" section) shows that the anatomically realistic VT of Maeda's model and in vivo VTs of live humans have the same acoustic potential as the n-tube models used to develop the MAS, and both the MAS and the MVS can generate the full F1-F2 vowel space. Furthermore, Fig. 6 also shows that the MVS is compatible with formant values obtained from a large set of human vocalizations for adult male speakers of eleven languages: Chinese, Dutch, American English, French, German, Japanese, Indian Malay, Brazilian and European Portuguese, Sardinian, and Swedish. However, crucially, Maeda's model generates only anatomically realistic articulations. In

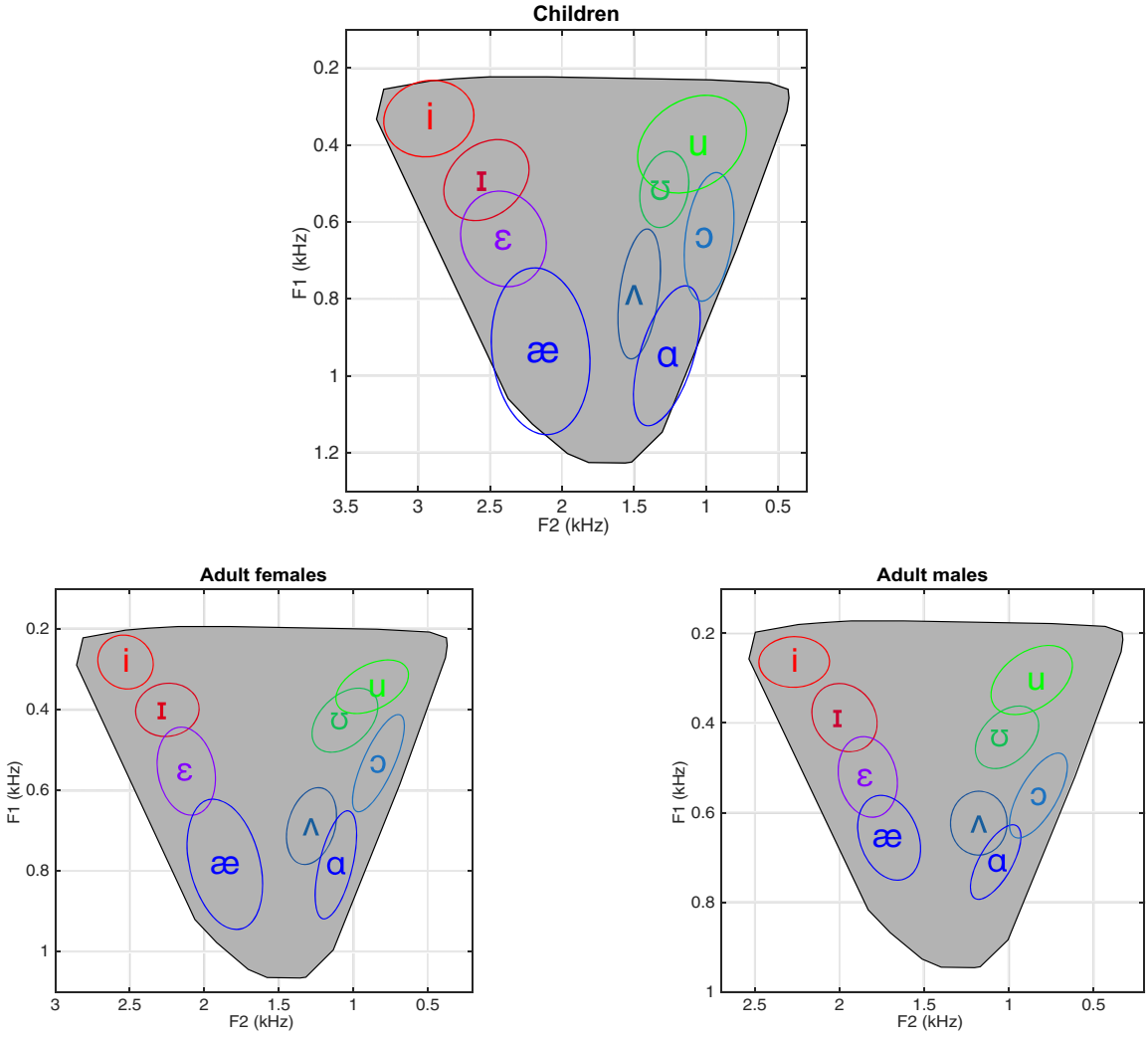

Fig. 4. American English vowels displayed within the MASs calculated by the four-tube model. Dispersion ellipses are for standard vowel values from Peterson and Barney (*12*) for young speakers, adult females, and adult males, displayed in a MAS for the appropriate VTL, 12.5, 15, and 17.5 cm, respectively.

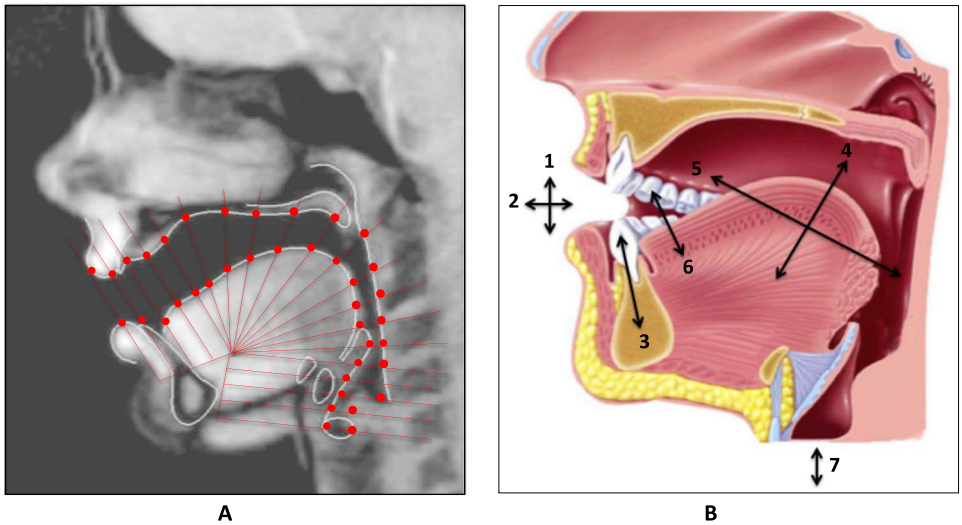

Fig. 5. Anthropomorphic articulatory model. (**A**) Measurement of the VT in the sagittal section with an analysis grid to sample the VT area in (typically) 30 planes transecting the VT. (**B**) Articulatory command parameters extracted using a principal components analysis of the sagittal section data. These parameters are interpretable with regard to the articulators: for the lips, (1) opening height and (2) protrusion; for the jaw, (3) the opening; for the tongue, the movements of (4) the dorsum, (5) the body, and (6) the apex; and for the larynx, (7) its height.

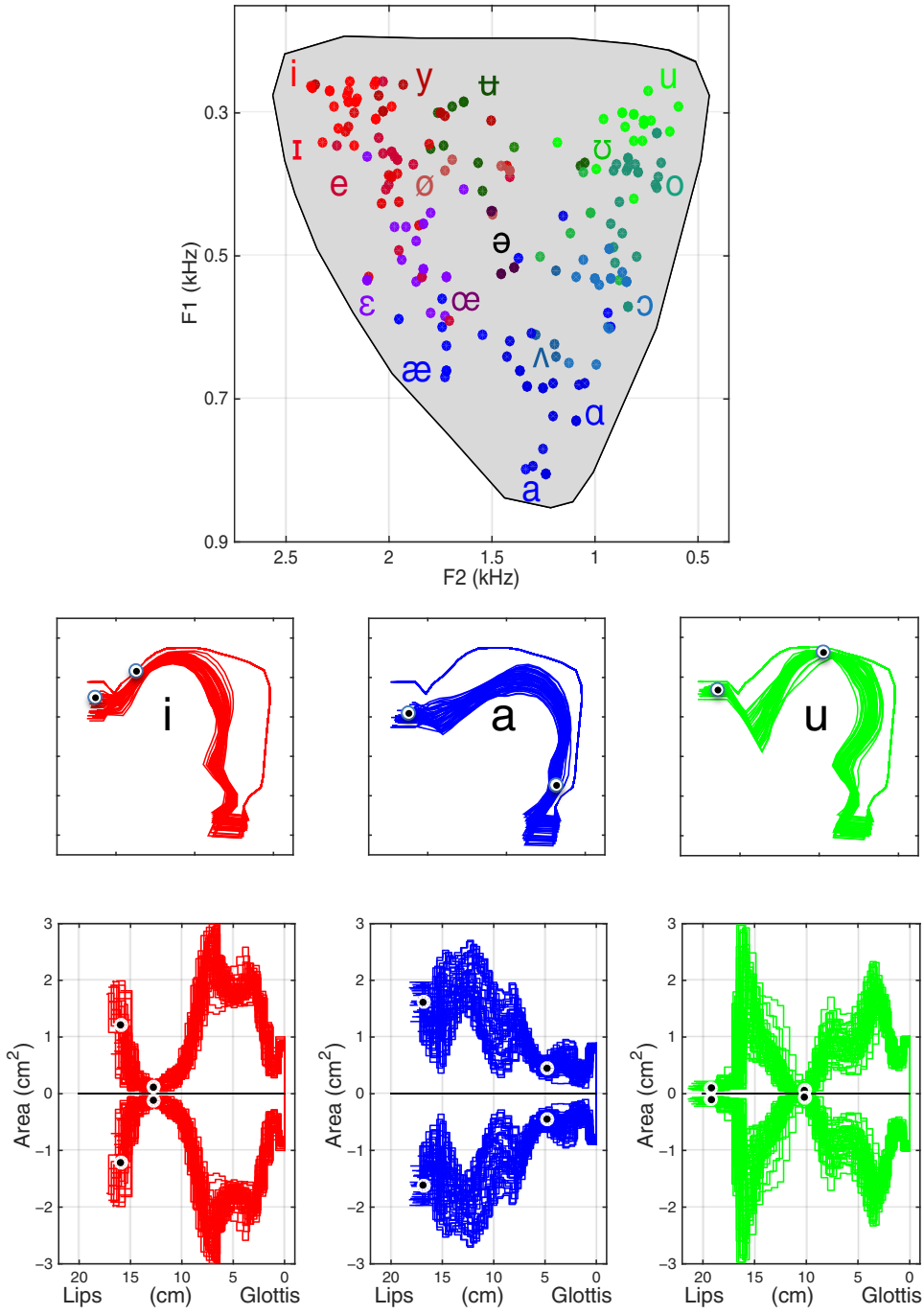

Fig. 6. MVS for the anthropomorphic articulatory model. (**Top**) Vowels of 11 world languages are projected into the maximal vowel space: Chinese (*172*), Dutch (*173*), American English (*12*, *87*, *174*), French (*175*, *176*), German (*177*), Japanese (*178*), Indian Malay (*179*), Brazilian and European Portuguese (*180*), Sardinian (*181*), and Swedish (*182*, *183*). Vowel labels from the original publications are coded according to the colors in Fig. 3. In addition, the low vowels, /æ a ɑ ɒ/, are grouped in a single macro-class. This F1-F2 space was generated by Maeda's model, and the model is validated by the correct placement of the different languages' vowels inside the MVS. (**Bottom**) Fifty sagittal sections (second row) and 50 area functions (third row) for /i a u/ (left to right) were obtained by acoustic-to-articulatory inversion from (F1, F2) values from French. Both are variations around those presented in Fig. 2, with dots in the sagittal sections and the area functions showing the positions of the labial and lingual constrictions, which thus allows computation of mean values for the crucial variables Xc, Ac, and Al. They illustrate, here for an adult male, prototypical VT forms, as well as the differing sensitivities noted in the "Tubes, cavities, constrictions, and acoustic resonances" section for lip area, Al, and tongue constriction position and area, Xc, and Ac. With this method, one can use formant values to generate plausible sagittal sections for all the vowels of the IPA.

a sense, it filters out the unrealistic articulations allowed within the MAS, and it has therefore been used, as *n*-tube models cannot, for inversion, to derive accurate articulatory configurations from vowels recorded in vivo (see Fig. 6, bottom).

VT growth and acoustic normalization

The anthropomorphic model introduced above (*111*) deals with adult anatomy and articulation, but VT growth is, of course, crucial for LDT. A number of studies have been done in this domain and led

to modifications of the articulatory model for dealing with the VT shape and size variations associated with ontogeny.

VT morphology depends on the bony structures of the head and cervical vertebrae. These situate and anchor the jaw, teeth, and hard palate, beyond which the tract is bounded by soft structures (tongue, pharyngeal walls, velum, and lips) with insertions into those bony structures. Over the course of ontogeny, the shape of the VT will therefore depend crucially on the growth of the skull and cervical vertebrae and on the positioning of the hyoid bone from which the larynx hangs. While modeling had previously been based principally on radiography of adult VTs, Goldstein (*26*) took on the task of integrating a wide array of anatomical data into an articulatory model of VT growth. From a

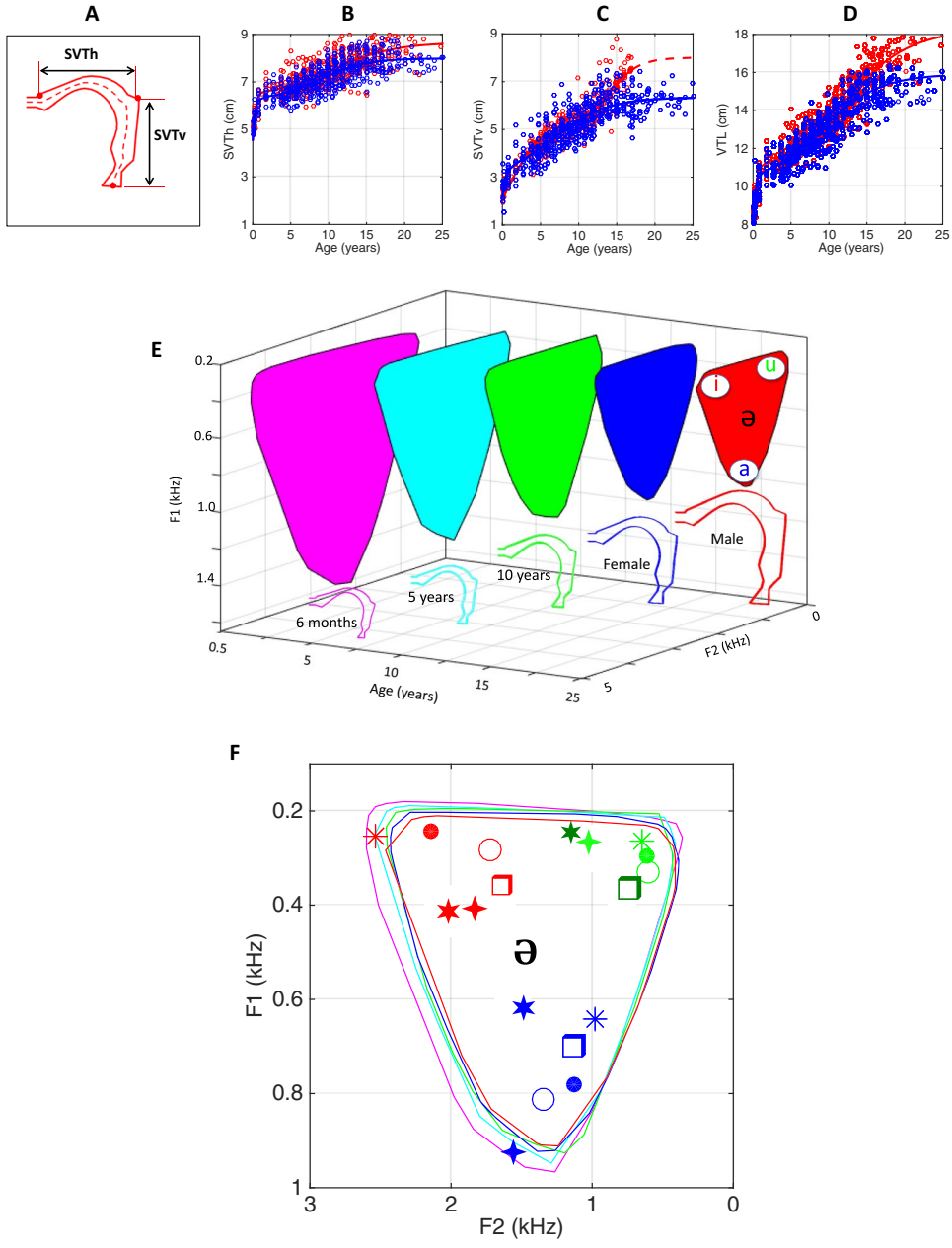

Fig. 7. Human VT growth. (**A**) Schematic VT illustrating the three pertinent measures: SVTh and SVTv, and the dashed median line where VTL is measured. (**B** to **D**) SVTh (B), SVTv (C), and VTL (D), from around birth to 25 years, females in red and males in blue. Results from Barbier *et al.* (*123*, *124*), as measured from four longitudinal databases, from 1 month to adulthood, using images from the American Association of Orthodontists (68 subjects, 966 x-rays), plus 12 fetal images from Montpellier University are shown. The data are optimized by two double logistic functions to account for growth in two phases, from birth to puberty and then from puberty to adulthood (*26*, *184*). Radiography did not capture the male glottis beyond 15 years, so the function in that range (dotted line) is an estimate. (**E**) VTs generated by VLAM (cf. VT growth and acoustic normalization section), from a human newborn through an adult male, along with their corresponding MVSs, with the three point vowels /i a u/, plus schwa /ə/. (**F**) Color-keyed boundaries of the five MVSs from (E) normalized to 17.5 cm, the mean length for a male at 25 years, with the /i a u/ vowels normalized from: predictions from Goldstein's model (*26*) for newborns ●●●, the extreme values from an imitation test for infants at 20 weeks (*137*) ✶✶✶, infants at 26 weeks (*126*) ✦✦✦, infants at 28 weeks (*129*) □□□, infants at 40 weeks (*128*) ○○○, and infants at 66 weeks (*127*) ✳✳✳ (plus schwa /ə/ for reference).

database she assembled from data on the bony anatomy in published medical literature based on radiography of the head and neck, she selected 19 measurements (16 distances and 3 angles) showing the changes in VT morphology from birth to adulthood for both sexes. In particular, her data integrate the uneven growth rates of the two key LDT parameters determining the VTL: the SVTh increase is small and slow while the SVTv increase is large and rapid (Fig. 7, B and C). All these data were fitted to age with simple or double logistic functions, which have been extensively studied and applied to a wide range of biological systems.

Elaborating on Mermelstein's adult model (*120*), Goldstein (*26*) used her rich dataset to develop a model of VT growth in the sagittal section, giving as output both the VT's area functions and the resulting formant values. Through vowel simulations placed (after VTL normalization) in the Peterson and Barney vowel space, she showed that as of birth:

> [T]he newborn is capable of producing /i/, /ɑ/, /u/, whereas the estimates of Lieberman would indicate that it is not. (*26*)

In our laboratory, we developed the variable linear articulatory model (VLAM) (*27*, *121*, *122*) to simulate VT growth from birth to adulthood based on the data published by Goldstein. This model uses a piecewise linear scaling to extend the anthropomorphic model developed by Maeda (*111*), which provided the adult settings, with scaling driven by age. VTL growth patterns were introduced via two scaling factors—one for the anterior oral cavity and the other for the posterior pharyngeal cavity—and an interpolation between them for the intermediate zone. The key elements for the model are Goldstein's simple or double logistic functions for age, conversion coefficients to pass from the sagittal section to the area function (*98*), and data from Barbier *et al.* (*123*, *124*) on VTL growth.

The MVSs simulated by VLAM at various ages (see Fig. 7E) show that the acoustic space is larger—when measured in hertz—for infants than for adults. This is a straightforward consequence of the fact that formants vary as the inverse of VTL. For example, a newborn with a VT half the length of an adult would potentially produce formants at twice the adult frequencies. This, in turn, implies that VTL is an appropriate normalization factor for comparing two vowel spaces using their formant frequencies as produced by VTs of different lengths [originally proposed by Mol (*125*); observational validation by Lee *et al.* (*87*)]. To do this, we simply normalize the formants (F) by taking account of the respective VTLs, l_1 and l_2

$$F\ell_2 = (\ell_1/\ell_2)F\ell_1$$

Thus, to compare formant frequencies of the same vowels as produced by VTs of different lengths, we take into account the inverse relationship of the lengths of the two VTs. This normalization process can be generalized to rescale whole F1-F2 vowel spaces, so that spaces, rather than specific vowels, can be compared. Doing so, we find that after normalization to a standard VTL, the potential MVSs for VTs across the human life span, from birth to adulthood, are all the same size (Fig. 7F).

Our model thus predicts that children are anatomically capable of the same vowel space as adults, and this is confirmed by an overwhelming amount of ground data collected on the vowel productions of infants and children (*122*, *126*–*136*). Notably, Kuhl and Meltzoff (*137*) show that infants at 20 weeks produce contrasting vowels when imitating adult /i a u/. Although the means of the distributions show that the vowels are phonetically "reduced" (i.e., closer to the schwa, /ə/, than the target to be imitated), their most extreme productions are definitely outside the schwa range, and for /a u/, they are (after normalization) within acceptable adult ranges (Fig. 7F).

These studies show that infants and children achieve contrasting, and even adult-like, productions starting well before developmental LD, even while their pharyngeal cavity is less than half its adult length. Responding to de Boer's defense of LDT's requirement of a lowered larynx based on an oversimplified model (*138*), we showed that the presence and action of the lips enable production of a normal F1-F2 space even in case of a high larynx (*139*). Overall, it appears that, regardless of the pharynx length, the size and shape of the MVS are quite stable from birth to adulthood (*29*) and that normalization, as exemplified in Fig. 7, is entirely justified. That is, articulatory modeling of the VT across human ontogeny shows that the vowel space is effectively equivalent before, during, and after LD and thus that LD and SVTh/SVTv ratio are irrelevant for the production of the full inventory of potential human vowel qualities.

UNIFORM TUBES CANNOT EXPLAIN PRIMATE VOCALIZATIONS: THE ACOUSTIC CAPACITIES OF NONHUMAN PRIMATES

Despite the claim by the LDT that nonhuman primates are limited to quasiuniform VT shapes and are thus incapable of modifying formants to contrast vocalizations, an increasing number of studies have come to differing conclusions. To show why, we must first explain the acoustics of uniform tubes and how departures from that configuration can be detected in nonhuman vocalizations. Then, we will show how a large amount of convergent data demonstrate that nonhuman primates do vary VT configurations and produce largely contrastive formant patterns, distinct from those of a uniform tube.

Acoustics of the uniform tube

Uniform tube: The ultimate simplification of VT models

The uniform tube is a key element for analyzing vocalizations in LDT (Fig. 8). A VT has variations of form along its length but can be configured in speech to maintain a generally uniform cross-sectional area from one end to the other, from the glottis to the lips. The uniform area allows it to be modeled by an acoustically equivalent cylindrical tube of the same length, closed at one end (the glottis) and open at the other (the lips). This is termed a quarter-wave resonator. The characteristics of such a model include that the formants are evenly spaced throughout the spectrum. Like all formants, their frequencies are proportional to the length of the tube, while the even spacing is diagnostic of a tube's uniformity. For our purpose, this means that for primates vocalizing through a uniform tube, those with long VTs will have formants evenly and closely spaced, while those with short VTs will have them evenly and distantly spaced. A further notable characteristic of uniform tubes closed at one end and open at the other is that the frequencies of higher formants are odd multiples of F1 (F2 = 3F1, F3 = 5F1, …) (Fig. 8).

Regarding the mathematical details, it can be shown that in a uniform tube closed at one end and open at the other, formant frequencies are defined as

$$Fi = (2i - 1)\, c/\lambda$$

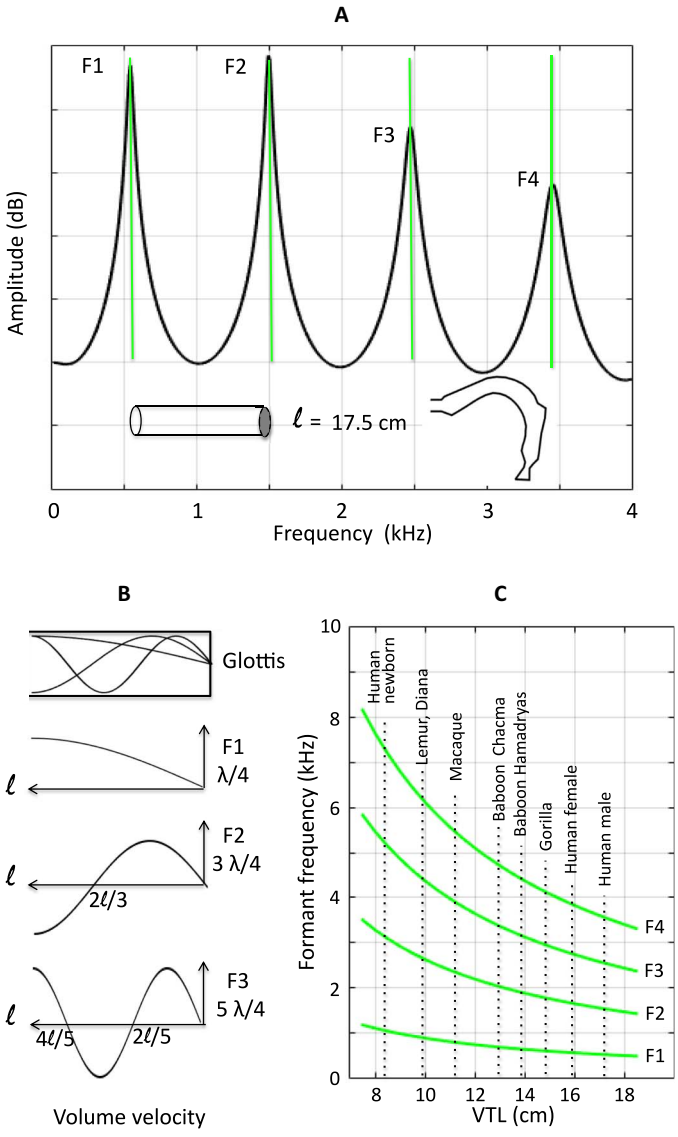

Fig. 8. Spectral properties of uniform tubes. (**A**) Acoustic transfer function of a uniform tube where l = 17.5 cm. Its calculation incorporates effects (lip radiation, wall losses, and viscosity) that slightly modify the theoretical formant values (0.5, 1.5, 2.5 kHz, …) (*97*, *185*). The human VT configured as a uniform tube produces the schwa vowel, /ə/. (**B**) Wave patterns in the uniform tube. The uniform tube models the VT as a quarter-wave resonator closed at the glottis and open at the lips, so tube acoustics generate formant values at frequencies defined by the odd multiples (1, 3, 5, 7, …) of one quarter of a wavelength, λ, equal to four times the length of the tube. The top of this panel shows the volume velocity wave along the tube for all of the first three formants. Below are the individual wave shapes for each of the first three formants, plotted individually (*14*). (**C**) Formant values for pertinent lengths of uniform tubes. This graph shows the formant values for uniform tubes with lengths across a range of VTLs, calculated (in kHz) as Fi = 35(2*i* − 1)/4ℓ, with lines marked at VTL values known for selected species noted in this paper. These are the formant values that should be expected when a VT of that length is configured as a uniform tube.

where Fi is the ith formant, i is an integer, c is the speed of sound (typically 350 m/s in the air), and λ is the wavelength of the sound, which is defined, for a uniform tube, as

$$\lambda = 4\ell$$

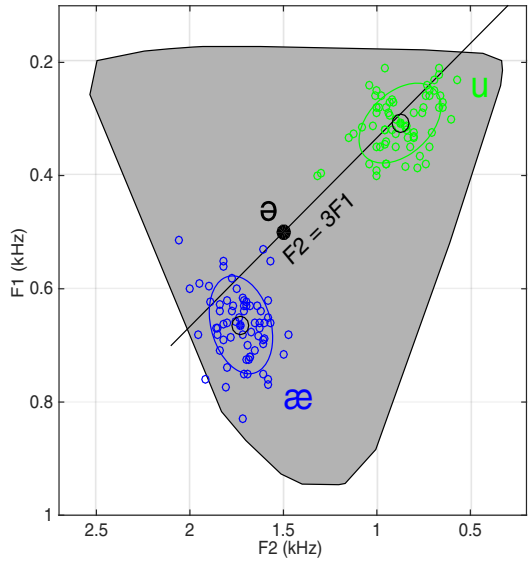

Fig. 9. F1-F2 MAS plane for men; dispersion ellipses for /u/, /æ/; and the straight line F2 = 3F1 produced by a uniform tube varying across VTL values. MAS set to VTL = 17.5 cm, data from Peterson and Barney (*12*). Points and means correspond to the values for adult males of the formants of /u/ and /æ/ and to the schwa-like /ə/. This shows that unless the VTL is known, the F2 = 3F1 criterion is insufficient for detecting the schwa-like productions of a uniform tube.

where l is the tube length (VTL for our purposes). To get F in kilohertz while supplying l in centimeters, the formula works out to

$$Fi_{\mathrm{kHz}} = (2i - 1) * 35/(4 * \ell_{\mathrm{cm}})$$

Formant pattern of the uniform tube as a test

Riede *et al.* (*88*) were the first to test animal vocalizations (namely, Diana monkey alarm calls) for compatibility with uniform tube characteristics, specifically with F2 = 3F1. They used a speech science approach to situate the calls by their formant patterns relative to human vowels produced by children with a similar VTL [as documented by (*87*)]. Riede and colleagues projected the formants into F1-F2 space to see whether they fell along the F2 = 3F1 line (Fig. 9). They showed that Diana monkey eagle alarm and leopard alarm calls are not schwa-like but are similar to /a/ as pronounced by 10- to 12-year-old children.

Note that the (F2 = 3F1) criterion is deceptive, as it underdetects certain nonuniform VT configurations as uniform. The F2 = 3F1 line crosses the entire vowel triangle and thus necessarily crosses a limited but important selection of nonuniform vowel configurations, from /u/ at one extreme to /æ/ at the other. Because the formants of baboon grunts fall along that line, but in the /u/ area, it is not surprising that they were mistakenly ascribed to production by a uniform tube [e.g., (*42*, *44*)]. Ultimately, the F2 = 3F1 criterion is necessary, but not sufficient, to detect a uniform tube. Information about the VTL is also needed to normalize formant values and locate them in an MVS relative to IPA vowels.

VTL estimation from formant patterns

As noted above, an estimate of VTL is crucial to represent vocalizations in an F1-F2 acoustic space and label them relative to IPA vowels. Because the formants of a uniform tube are evenly spaced through the

frequency spectrum, the VTL can be found directly from any given formant (*Fi*), using

$$\ell = (2i - 1)c/4Fi$$

where *c* is 350 m/s, the speed of sound in humid VT air at 35°C (approximating primate body—and VT—temperature). We therefore propose that all VTL estimates from all formants using this method should agree for the tube to be recognized as uniform with the same resulting *l*. Otherwise, the tube is nonuniform, and some other method must be found to estimate the VTL.

Using this test, we have detected various important discrepancies, for instance, regarding baboon grunts observed by Rendall and colleagues (*71*). In a uniform tube, the estimations using the first four formants (F1, …, F4) measured in males would imply VTLs ranging from 23.3 to 32.6 cm and 16.3 to 22.9 cm for females. These are seriously disparate VTL values and entirely implausible for baboons, where we have measured VTLs of 13.5 cm for males and 11 cm for females (*35*). Similarly, Owren and colleagues found means of VTLs estimated from the first four formants (F1, …, F4) varying from 15.9 to 19.7 cm. These discrepancies intrigued the researchers, who noted:

> [T]he derived lengths were significantly longer and shorter, respectively, than expected (based on the mean overall estimate [17.9 cm]). (*44*)

Note that in these examples, the VTL is frequently overestimated, as it would be in human speech if calculated from the first formants of /u/. For instance, the F1 and F2 means found by Peterson and Barney for American English /u/ in adult males would, respectively, imply VTLs of 29.2 and 30.3 cm, too long—absurdly so—for a VTL often cited at 17.5 cm, but analyzed as 19.6 cm for /u/ (*140*). The pertinent vocalizations, baboon grunts, do show /u/-like formant patterns, as we will show below.

An estimate based on a single formant is necessarily imprecise, considering the intrinsic uncertainty associated with formant estimation. Hence, various methods have been proposed capitalizing on the previous equation but adding statistical tools to make the evaluation more robust.

One method that has been proposed is

$$\ell = c/2\mathrm{Df}$$

where Df is the mean of several successive intervals from F1 to Fn (*43*), which reduces to

$$\mathrm{Df} = (\mathrm{F}n\text{–}\mathrm{F1})/(n\text{–}1)$$

A second, more sophisticated and precise tool uses the slope Df of the linear regression between the formant numbers (1, 2, 3 … *n*) and their respective values to estimate formant spacing (*141*).

With $n > 4$, these two propositions have the advantage of minimizing the importance of F1 and F2, the formants most influenced by VT deviations from a uniform tube configuration, and they tend to show that higher formants give good VTL estimates even for nonuniform VTs. Expressed differently, the frequencies of higher formants, regardless of the VT's shape, tend to converge on those of a uniform tube. If the higher formants are detectable, as is often the case with primate vocalizations, these two methods are effective for estimating VTL, even in the case of nonuniform tube configurations. In the following sections, we have relied on the second method as potentially more robust.

LDT falsified in nonhuman primates

With this battery of principles and tools, we can now address the existing acoustic data on primate vocalizations, and we can plan and execute targeted experiments and discuss their implications for the LDT.

Discrepancies detected

A number of papers in the past 15 years have led their authors to raise doubts that their acoustic measurements of nonhuman primate vocalizations were compatible with the even formant spacing expected from a

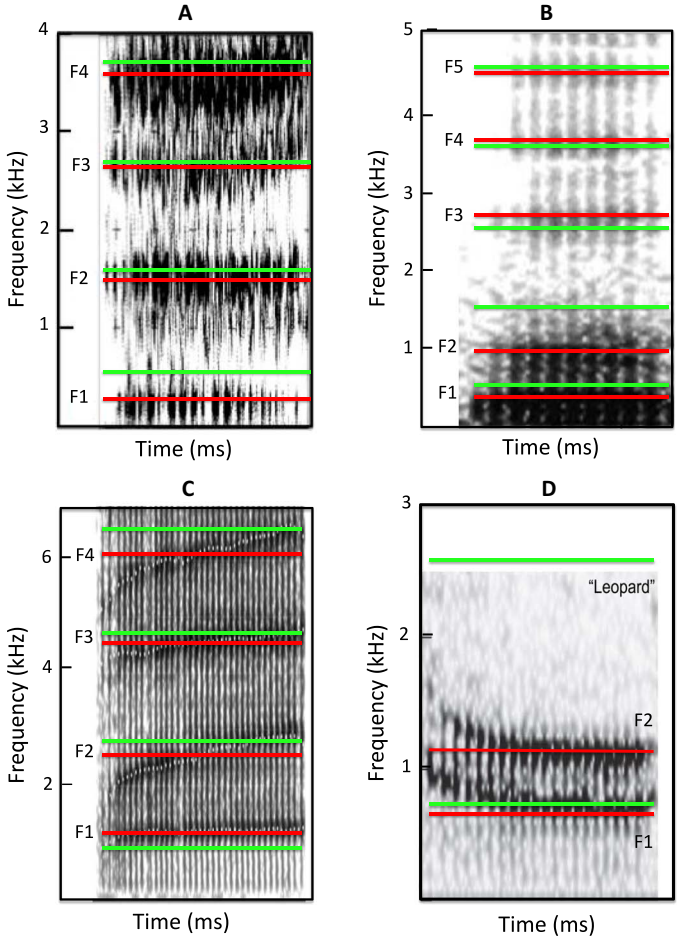

Fig. 10. Formant patterns of vocalizations (red lines) diverging from those of uniform tube (green lines). The green lines represent calculated estimates of the expected formants from a uniform tube of the appropriate VTL, while the red lines represent either values reported by the authors or means calculated from their figures, adapted here for graphic purposes. VTLs for (A) to (C) were estimated with the Reby and McComb method using five formants for (A) and (B) and six for (C). VTL for (D) was measured from x-rays. (**A**) Double grunt of *Gorilla gorilla beringei* [Fig. 1 of (*41*)]; VTL = 16.4 cm. F1 is much too low for a uniform tube. (**B**) Grunt of *Papio hamadryas* adult female [Fig. 2 of (*102*)]; VTL = 17.3 cm. F2 is much too low for a uniform tube. (**C**) Roar of *Eulemur mongoz* [Fig. 1A of (*186*)] with initial formant transition; VTL = 10.7 cm. Final formant values are compatible with a uniform tube, but the variations of F2, F3, and F4, while F1 stays stable, are not compatible with a uniform tube. (**D**) Leopard alarm call of male Diana monkey [Fig. 2 of (*88*)]; VTL = 10 cm. As noted by the authors themselves, neither the F2 values nor the (F1, F2) variations along the trajectory are compatible with a uniform tube.

uniform tube. The literature contains numerous remarks by primatologists confronted by formant patterns unexplained by LDT, such as:

> For many years, it has been the default assumption that mammalian vocal tracts, including those of non-human primates, resemble a uniform or flared tube during vocalization (Lieberman, 1968; Lieberman et al., 1969; Shipley et al., 1991). In a uniform cylindrical vocal tract the resonance frequencies are expected to appear as odd numbered multiples of the first resonance, and all resonances are evenly spaced. … More recently, various studies challenged this view, by suggesting that some animal vocalisations are the product of non-uniform vocal tracts (e.g., Owren et al., 1997). For example, we have previously demonstrated that the location of the first (F1) and second (F2) formant in Diana monkey alarm calls cannot be explained by a uniform vocal tract but must be the result of a more complex vocal tract geometry. (*88*)

We show in Fig. 10 a selection of spectrograms with distinctly irregular formant spacing, which thus indicate a departure from the schwa-like acoustics resulting from a VT shaped like a uniform tube. Figure 10 (C and D) also shows modification of the VT configuration over the course of the vocalization. Riede *et al.* (*88*) suggest that "the acoustic structure of these calls is the product of a non-uniform vocal tract capable of some degree of articulation". Pisanski and colleagues (*5*) "suggest that this may represent a living relic of early vocal control abilities that led to articulated human speech".

Two focused experiments, two explicit refutations

Two articles appearing 3 weeks apart (*35*, *142*) ultimately proved the LDT untenable. In the first article, Fitch *et al.* showed that the macaque VT attained articulatory configurations distinctly different from a uniform tube during vocalization and also during feeding and facial expressions. The authors conclude:

> We demonstrate that the macaque vocal tract could easily produce an adequate range of speech sounds to support spoken language, showing that previous techniques [Lieberman, Klatt, Wilson, 1969] based on postmortem samples drastically underestimated primate vocal capabilities. Our findings imply that the evolution of human speech capabilities required neural changes rather than modifications of vocal anatomy. Macaques have a speech-ready vocal tract but lack a speech-ready brain to control it. (*142*)

One of the authors later states that:

> Now nobody can say that it's something about the vocal anatomy that keeps monkeys from being able to speak. [Ghazanfar in (*143*)]

After years of sometimes harsh debate, during which they essentially reversed their position (*138*, *144*, *145*), they finally converge on what they term the neural hypothesis, a conclusion advanced by Boë over a decade earlier based on simulations of a VT with a small pharynx:

> Endowed with a small pharyngeal cavity, monkeys exhibit the same vocal tract configuration as newly-born infants, but if they do not produce vowels, it is not due to this resemblance. ... No, if monkeys do not talk, according to present evidence, this is due to a lack of appropriate cortical equipment. (*27*)

For their part, Boë and colleagues show (*35*) that baboons [*Papio papio* (*146*)] naturally produce vowel-like sounds sharing the (F1, F2) formant structure of the human [i æ ɑ ɔ u] vowels, including the proto-bisyllabic call wahoo, and that those vocalic qualities are organized in a proto-system similar to that of humans:

> Our findings therefore reveal a loose parallel between human vowels and baboon VLSs [Vowel Like Segments] by demonstrating that both have a phonetic inventory of vocalic qualities differentiated by formant structure and that these structures are characteristic properties of vocalizations produced in distinct social contexts or for different functions. From an evolutionary standpoint, demonstration of a two [articulatory] axis vocalic proto-system in baboons suggests that the human vocalic system did not emerge de novo but originates from articulatory capacities already present in our common ancestors. (*35*)

After earlier challenges to LDT for children and Neanderthals, these two articles bring the challenge to the realm of nonhuman primates using two complementary approaches. Fitch and colleagues used medical imagery techniques to directly determine VTL and VT configurations, then articulatory-acoustic modeling to estimate the formants for a male rhesus macaque, and finally generated an example of hypothetical macaque articulatory capacities by dynamic synthesis of a whispered sentence. Boë and colleagues submitted more than a thousand naturally produced baboon vocalizations [recorded and ethologically documented from a research baboon troop in semi-liberty (*59*)] to LPC analysis for measurement and assignment to five phonetic categories corresponding to five ethological situations. They added an anatomical study to measure the VTL and document the tongue musculature recruited to produce the vocalizations. With the VTL normalized and a corresponding MAS generated, the superimposition of the Peterson and Barney vowels allowed them to categorize the baboon vocalizations as using five separate proto-vowels. In Fig. 11, we present a MAS with dispersion ellipses for selected Peterson and Barney vowels (from child talkers) for comparison with (i) the macaque findings of Lieberman *et al.* (*10*) and Fitch *et al.* (*142*) and (ii) our own 2017 findings concerning baboons. Unlike Lieberman *et al.*, both Fitch *et al.* and Boë *et al.* found articulations well beyond the schwa-like uniform tube configuration.

Note the enlargement of the acoustic space found for vocalizations over the three studies, which may be explicable by the different conditions of data collection for each study. Lieberman and colleagues used molds of a postmortem VT, which fixes the articulatory configuration and makes it difficult to challenge any extrapolations from the VT's form. Fitch *et al.* used x-ray video of live animals, but any radioimaging setup is necessarily constraining, and both the surroundings and the absence of companion troop members must leave the subject in an ethologically unnatural situation. The vocalizations used in the Boë *et al.* study were produced spontaneously by a troop of baboons living mainly outdoors in semi-liberty under near-natural conditions. Apparently, the more natural the data collected, the larger the acoustic space covered.

These two studies are functional replications of foundational LDT studies, Fitch *et al.* of (*10*) and Boë *et al.* of (*8*). Their findings constitute strong and probably definitive arguments for the refutation of the LDT.

Diamonds in the coal: New findings from vintage data

From pertinent publications, we have selected data regarding various nonhuman primate vocalizations showing irregularly spaced formant

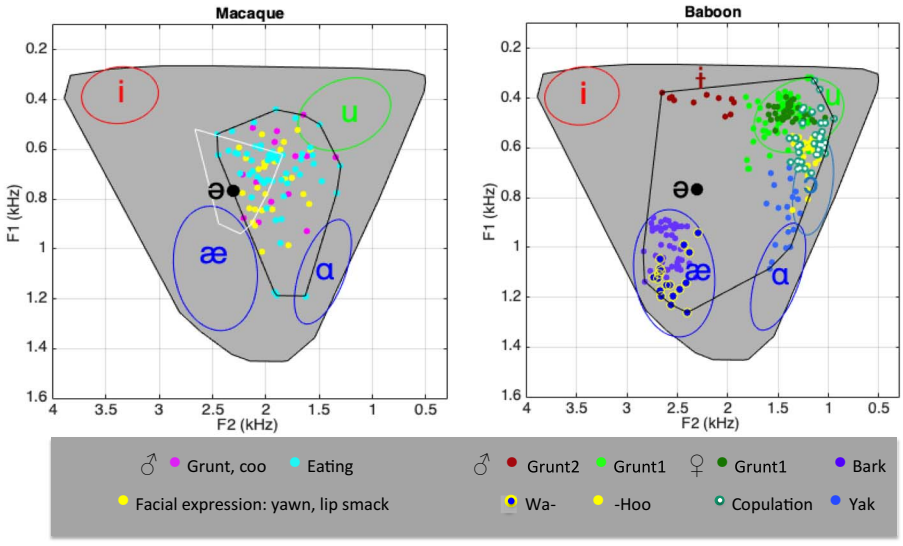

Fig. 11. Phonetic qualities of macaque and baboon articulations within the MAS. (**Left**) Macaque vowel spaces, according to Lieberman *et al.* (*10*) (white line) and according to Fitch *et al.* (*142*) with the convex hull (black line) enclosing data from vocalizations, facial expressions, and eating. (**Right**) Vocalizations from Boë *et al.* (*35*) with the convex hull (black line) enclosing data from vocalizations. For both sides, all the data were either obtained from or normalized to a reference VT of 11.4 cm [the VTL of the macaque in (*142*)] and presented along with the dispersion ellipses (color-coded as in Fig. 3) for Peterson and Barney's data for children (*12*), also normalized to a VTL of 11.4 cm.

patterns impossible to produce with the uniform VT corresponding to schwa. We noted the published formant values and determined the associated VTLs, either measured by the authors using radiography or MRI or estimated by us from those measured formants using Reby and McComb's procedure (*141*). We then generated adjusted formant values, through normalization to a standard VT with a VTL of 11.4 cm, using our previously elaborated procedure (see 3.4.2 above). In Fig. 12, we have projected the normalized formant data into the corresponding F1-F2 space of the MAS, along with dispersion ellipses for selected vowels from Peterson and Barney's data for children's vowels (*12*).

Note that all utterances are situated inside the normalized MAS, which confirms the validity of this method. From these results, we can make several observations. First, we note that the selected vocalizations are not the schwa-like productions expected from a VT configured as a uniform tube but are instead dispersed in a large part of the MAS and show the various distinct formant patterns appropriate to /ɨ I æ o ʊ u/ vowel qualities. Second, there are no occurrences of the /ɑ/ quality that we had found in the baboon yak, although some vocalizations' formants place them between /æ/ and /ɑ/, in /a/ territory (as in French). Third, there are occurrences of /I/, immediate neighbor to the high front /i/ vowel quality, that had not previously been recognized in the literature as a vocalization from a living nonhuman primate. Fourth, while not our focus in this paper, we note that these different vowel qualities are produced by species whose hyoid bones are not shaped like human hyoids (e.g., baboons and macaques), strongly implying [contra, e.g., (*147*)] that hyoid shape is irrelevant for speech emergence.

Last, we conclude from this preponderance of appropriately VTL-normalized evidence that multiple primate species, from gorillas, through certain Old World monkeys, to even lemurs, produce non-schwa vowel qualities through their VT's deviations from a uniform tube configuration. We further conclude that these counterexamples refute the core claims of LDT: that pre-AMHS hominids can only produce schwa-like vowels, while AMHS adults alone can articulate the full array of vowels because of their descended larynx and large pharynx and the resulting 1:1 SVTh/SVTv ratio.

LESSONS LEARNED FROM THE LD CONTROVERSY

The LDT claimed that production of differentiated vowel qualities was anatomically impossible except in AMHS adults. Since its formulation and diffusion beginning in the 1970s, a hypothesis has arisen [e.g., see (*3*)] that the emergence of speech and language was recent, sudden, and simultaneous: recent because contrasting vowels (and hence phonology) only became possible in AMHS [whose emergence was recently pushed back from ~200 to ~300 ka ago; (*148*)]; sudden because 300 ka is extremely short on an evolutionary time scale; and simultaneous because both speech and language would need that short period to develop.

Three major facts emerge from our review of ground-truth data in light of articulatory-acoustic models and the use of the modern tools from speech. First, even among primates, LD is not uniquely human. Second, LD is not required to produce contrasting formant patterns in vocalizations. Third, nonhuman primates do produce vocalizations with contrasting formant patterns. Thus, the evidence is now overwhelming, from arguments based on both observations and modeling, that the LDT is no longer tenable. The larynx did descend in phylogeny (and still does in ontogeny), but the descent was neither necessary nor sufficient for the emergence of speech from primate vocalization. We find that the anatomical potential to produce and perceive sounds differentiated by their formants began at the latest by the time of our last common ancestor with Old World monkeys (Cercopithecoidea) about 27 Ma ago. That element of speech, at least, would have been available beginning then as the channel for communication by language during its emergence at any later time.

LDT was the most broadly disseminated and, to the best of our knowledge, the only broadly accepted claim to establish a time frame of

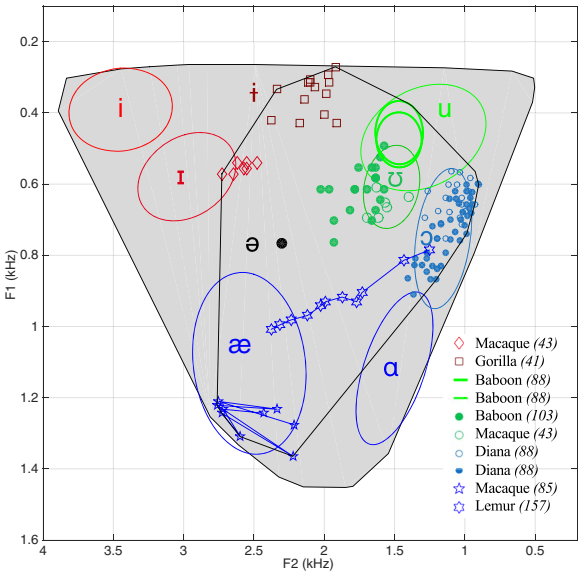

Fig. 12. Reframing vintage data in the normalized MAS. This figure collects and presents data from a variety of published articles on primate vocalizations. The MAS and all data are normalized to VTL = 11.4 cm [the VTL of the macaque in (*142*)]. Original (prenormalization) VTLs are determined using the articles' formant values and Reby and McComb's method (*141*) (discussed in the "VTL estimation from formant patterns" section) except VTLs for chacma baboon males: estimated from known VTLs of other baboon species; chacma baboon females: 25% shorter than males, because formants are 25% higher (*187*); and Diana monkeys: from radiography (*88*). Vowel dispersion ellipses for /i ɪ æ ɔ ʊ u/ are from Peterson and Barney's data for children (*12*). Reframed primate data, starting from /ɪ/ and proceeding clockwise, are as follows: threat calls from rhesus macaque (*Macaca mulatta*) subjects 73 to 91 (Fig. 4) (*43*); double grunts of mountain gorillas (*G. gorilla beringei*) [table II of (*41*)]; grunts of chacma baboon (*Papio hamadryas ursinus*) males (larger bold light green ellipse) and females (smaller bold light green ellipse) [Fig. 1B of (*88*), data collected for (*44*), first presented by (*187*)]; grunts of hamadryas baboon (*P. hamadryas*) subject SI (Fig. 3B) (*102*); threat calls from rhesus macaque (*M. mulatta*) subjects 241-95 (Fig. 4) (*43*); eagle (blue circle) and leopard (blue dot) alarm calls of male Diana monkeys (*Cercopithecus diana*) [Fig. 1A of (*88*), data collected for (*45*)] single exemplar of rhesus macaque (*M. mulatta*) girney, with lines connecting formant values affected by lip and jaw modulation (Fig. 1G) (*85*); single exemplar of mongoose lemur (*E. mongoz*) alarm long grunt, with lines connecting shifting formant values (Fig. 1A) (*186*). The convex hull (black line) encloses all the primate vocalizations shown, thus representing the collected vowel space of primates as documented to date, and is subject to enlargement in future studies.

the emergence of speech and language based on anatomical evidence. We have used the comparative method to show, using our own data and reanalyzed data from others, that a key element of human speech, contrasting vowel qualities, was available to our ancestor species at least 100 times earlier than previously theorized under LDT.

Of course, many other elements have been addressed in the lively discussions about the emergence of speech and spoken language, including such topics as cranial anatomy of premodern humans (*149*, *150*), auditory sensitivity changes (*151*), risk of choking (*152*), laryngeal motor control (*153*, *154*), and functional neuroanatomy (*155*, *156*) [see a recent review of some of these topics in (*157*)]. The broader problem of the origin of language per se engages a further, more abstract set of topics, involving the biological, cognitive, and social conditions of language emergence [e.g., see the special issue introduced by (*158*)]. We intentionally declined to engage with these topics because, as we stated in Introduction, the ability to produce contrasting vowels is a foundational concern for the evolution of speech and spoken language, and ultimately of language in general, since the development of other aspects of speech (e.g., consonants, syllables, speech perception, and phonology), and thereby of a spoken lexicon available for syntactic manipulation, depends at a minimum on prior mastery of the articulatory ability to transmit those contrasting vowels. Our finding that that ability must have arisen so much earlier in the primate line casts a new light on several other lines of evidence and argumentation.

Specifically, the idea of recent, sudden, and simultaneous emerging of speech and language is no longer plausible. The dawn of speech in the form of contrasting vowel sounds is not recent, but early. The full process of speech emergence was therefore not sudden but extended, probably occurring in stages about which we can now begin to theorize. The final developments of this long process doubtless coincided with the emergence of language, but conceiving that much shorter process as simultaneous with speech emergence as a whole is no longer warranted.

Thus, we believe that the present refutation of the LDT should have a profound and liberating effect on our understanding of human evolution because, without the time limit imposed by LD, a variety of other hypotheses about language emergence can now be entertained. While speech had been thought of as enabling communication of already developed linguistic cognition, should it now be thought of as an early driver of linguistic cognitive development? For instance, gestural theories of language origin (*33*, *159*) attracted increased support, in part, for allowing a longer window for language phylogeny from observed primate gesture, so should we suddenly abandon those theories and search for language phylogeny solely in observed calls, or should we investigate more broadly for communication by gesture and vocalization combined? In light of demonstrations that nonhuman primates can modify their VT shapes to differentiate vocalic qualities, should this be understood as exaptation of structures dedicated to breathing, chewing, and swallowing, and does this strengthen the "frame and content theory" argument (*160*, *161*) that syllabic and phonological structure arose by exaptation of control of those same processes? The dawn of syntax had been sought in some triggering neurocognitive event (*3*, *162*) contemporaneous with LD in AMHS, so should it now be imagined as a slower, more intricate exaptation of prior cognitive faculties common to primates? Other similar lines of reasoning, formerly understood as too "early" to apply to speech and language because they concern behavior of living primates, must also be reevaluated as possibly less premature: turn-taking (*163*, *164*), laryngeal control (*5*, *165*), audience effects (*166*, *167*), and cultural transmission (*168*–*170*).

A good deal more comparative work with living primates is called for. We eagerly await exploratory reports from these newly reopened vistas.

REFERENCES AND NOTES

1. M. H. Christiansen, S. Kirby, in *Language Evolution*, M. H. Christiansen, S. Kirby, Eds. (Oxford Univ. Press, 2003), pp. 1–15.
2. B. Latour, *Science in Action: How to Follow Scientists and Engineers Through Society* (Harvard Univ. Press, 1987).
3. J. J. Bolhuis, I. Tattersall, N. Chomsky, R. C. Berwick, How could language have evolved? *PLOS Biol.* **12**, e1001934 (2014).
4. J.-J. Hublin, How to build a Neandertal. *Science* **344**, 1338–1339 (2014).
5. K. Pisanski, V. Cartei, C. McGettigan, J. Raine, D. Reby, Voice modulation: A window into the origins of human vocal control? *Trends Cogn. Sci.* **20**, 304–318 (2016).
6. F. Berthommier, L.-J. Boë, A. Meguerditchian, T. R. Sawallis, G. Captier, in *Origins of Human Language: Continuities and Discontinuities with Nonhuman Primates*, L.-J. Boë, J. Fagot, P. Perrier, J.-L. Schwartz, Eds. (Peter Lang, 2018), pp. 101–135.
7. W. N. Kellogg, Communication and language in the home-raised chimpanzee. *Science* **162**, 423–427 (1968).

8. P. Lieberman, Primate vocalizations and human linguistic ability. *J. Acoust. Soc. Am.* **44**, 1574–1584 (1968).
9. P. Lieberman, K. S. Harris, P. Wolff, L. H. Russel, Newborn infant cry and nonhuman primate vocalization. *J. Speech Hear. Res.* **14**, 718–727 (1971).
10. P. Lieberman, D. H. Klatt, W. H. Wilson, Vocal tract limitations on the vowel repertoires of rhesus monkey and other nonhuman primates. *Science* **164**, 1185–1187 (1969).
11. P. Lieberman, E. S. Crelin, On the speech of Neanderthal man. *Linguist. Inq.* **2**, 203–222 (1971).
12. G. E. Peterson, H. L. Barney, Control methods used in a study of the vowels. *J. Acoust. Soc. Am.* **24**, 175–184 (1952).
13. P. Passy, Our revised alphabet. *Phon. Teach.* **1888**, 57–60 (1888).
14. T. Chiba, M. Kajiyama, *The Vowel: Its Nature and Structure* (Phonetic Society of Japan, 1941).
15. I. Maddieson, in *Patterns of Sounds* (Cambridge Univ. Press, 1984).
16. V. Negus, *The Comparative Anatomy and Physiology of the Larynx* (Heinemann, 1949).
17. W. T. Fitch, J. Giedd, Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *J. Acoust. Soc. Am.* **106**, 1511–1522 (1999).
18. D. E. Lieberman, R. C. McCarthy, The ontogeny of cranial base angulation in humans and chimpanzees and its implications for reconstructing pharyngeal dimensions. *J. Hum. Evol.* **36**, 487–517 (1999).
19. P. Lieberman, Current views on Neanderthal speech capabilities: A reply to Boe et al. (2002). *J. Phon.* **35**, 552–563 (2007).
20. P. Lieberman, The evolution of human speech: Its anatomical and neural bases. *Curr. Anthropol.* **48**, 39–66 (2007).
21. P. Lieberman, The evolution of language and thought. *J. Anthropol. Sci.* **94**, 127–146 (2016).
22. S. J. Gould, E. S. Vrba, Exaptation-A missing term in the science of form. *Paleobiology* **8**, 4–15 (1982).
23. D. Falk, Comparative anatomy of the larynx in man and the chimpanzee: Implications for language in Neanderthal. *Am. J. Phys. Anthropol.* **43**, 123–132 (1975).
24. E. L. Du Brul, Biomechanics of speech sounds. *Ann. N. Y. Acad. Sci.* **280**, 631–642 (1976).
25. P. Houghton, Neandertal supralaryngeal vocal tract. *Am. J. Phys. Anthropol.* **90**, 139–146 (1993).
26. U. G. Goldstein, thesis, Massachusetts Institute of Technology (1980).
27. L.-J. Boë, in *Proceedings of the 14th International Congress of Phonetic Sciences*, J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, A. C. Bailey, Eds. (University of California, 1999), vol. 3, pp. 2501–2504.
28. L.-J. Boë, in *Percevoir: Monde et Langage. Invariance et Variabilite du Sens vecu*, D. Keller, J.-P. Durafour, J.-F. Bonnot, R. Sock, Eds. (Editions Mardaga, 2001).
29. L.-J. Boë, P. Badin, L. Ménard, G. Captier, B. Davis, P. MacNeilage, T. R. Sawallis, J.-L. Schwartz, Anatomy and control of the developing human vocal tract: A response to Lieberman. *J. Phon.* **41**, 379–392 (2013).
30. D. Sperber, in *Mapping the Mind: Domain Specificity in Cognition and Culture*, L. A. Hirschfeld, S. A. Gelman, Eds. (Cambridge Univ. Press, 1994), pp. 39–67.
31. D. Sperber, *La Contagion des Idées* (Odile Jacob, 1996).
32. G. W. Hewes, Primate communication and the gestural origin of language. *Curr. Anthropol.* **14**, 5–24 (1973).
33. M. C. Corballis, The gestural origins of language: Human language may have evolved from manual gestures, which survive today as a "behavioral fossil" coupled to speech. *Am. Sci.* **87**, 138–145 (1999).
34. M. A. Arbib, K. Liebal, S. Pika, Primate vocalization, gesture, and the evolution of human language. *Curr. Anthropol.* **49**, 1053–1076 (2008).
35. L.-J. Boë, F. Berthommier, T. Legou, G. Captier, C. Kemp, T. R. Sawallis, Y. Becker, A. Rey, J. Fagot, Evidence of a vocalic proto-system in the baboon (*Papio papio*) suggests pre-hominin speech precursors. *PLOS ONE* **12**, e0169321 (2017).
36. L.-J. Boë, T. R. Sawallis, J. Fagot, F. Berthommier, in *Origins of Human Language: Continuities and Discontinuities with Nonhuman Primates*, L.-J. Boë, J. Fagot, P. Perrier, J.-L. Schwartz, Eds. (Peter Lang, 2018), pp. 59–74.
37. B. S. Atal, The history of linear prediction. *IEEE Signal Process. Mag.* **23**, 154–161 (2006).
38. R. J. Andrew, in *Origins and Evolution of Language and Speech,* S. R. Harnad, H. D. Steklis, J. Lancaster, Eds. (New York Academy of Sciences, 1976), pp. 673–693.
39. B. Richman, Some vocal distinctive features used by gelada monkeys. *J. Acoust. Soc. Am.* **60**, 718–724 (1976).
40. M. J. Owren, R. H. Bernacki, The acoustic features of vervet monkey alarm calls. *J. Acoust. Soc. Am.* **83**, 1927–1935 (1988).
41. R. M. Seyfarth, D. L. Cheney, A. H. Harcourt, K. J. Stewart, The acoustic features of gorilla double grunts and their relation to behavior. *Am. J. Primatol.* **33**, 31–50 (1994).
42. D. Rendall, M. J. Owren, P. S. Rodman, The role of vocal tract filtering in identity cueing in rhesus monkey (*Macaca mulatta*) vocalizations. *J. Acoust. Soc. Am.* **103**, 602–614 (1998).
43. W. T. Fitch, Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *J. Acoust. Soc. Am.* **102**, 1213–1222 (1997).
44. M. J. Owren, R. M. Seyfarth, D. L. Cheney, The acoustic features of vowel-like grunt calls in chacma baboons (*Papio cyncephalus ursinus*): Implications for production processes and functions. *J. Acoust. Soc. Am.* **101**, 2951–2963 (1997).
45. T. Riede, K. Zuberbühler, The relationship between acoustic structure and semantic information in Diana monkey alarm vocalization. *J. Acoust. Soc. Am.* **114**, 1132–1142 (2003).
46. J. J. Ohala, An ethological perspective on common cross-language utilization of F_0 of voice. *Phonetica* **41**, 1–16 (1984).
47. J. J. Ohala, in *Proceedings of the 4th Seoul International Conference on Linguistics* (Linguistic Society of Korea, 1997), pp. 98–103.
48. P. Marler, in *Primate Behavior: Field Studies of Monkeys and Apes*, I. DeVore, Ed. (Holt, Rinehart and Winston, 1965), pp. 544–584.
49. P. Marler, in *Progress in Ape Research*, G. H. Bourne, Ed. (Academic Press, 1977), pp. 85–96.
50. J. Fischer, Primate vocal production and the riddle of language evolution. *Psychon. Bull. Rev.* **24**, 72–78 (2017).
51. T. T. Struhsaker, in *Social Communication Among Primates*, S. A. Altmann, Ed. (University of Chicago Press, 1967), pp. 281–324.
52. R. M. Seyfarth, D. L. Cheney, The assessment by vervet monkeys of their own and another species' alarm calls. *Anim. Behav.* **40**, 754–764 (1990).
53. M. J. Owren, Acoustic classification of alarm calls by vervet monkeys (*Cercopithecus aethiops*) and humans (*Homo sapiens*): I. Natural calls. *J. Comp. Psychol.* **104**, 20–28 (1990).
54. M. J. Owren, Acoustic classification of alarm calls by vervet monkeys (*Cercopithecus aethiops*) and humans (*Homo sapiens*): II. Synthetic calls. *J. Comp. Psychol.* **104**, 29–40 (1990).
55. R. M. Seyfarth, D. L. Cheney, The acoustic features of vervet monkey grunts. *J. Acoust. Soc. Am.* **75**, 1623–1628 (1984).
56. B. Schuller, S. Steidl, A. Batliner, F. Burkhardt, L. Devillers, C. Müller, S. Narayanan, Paralinguistics in speech and language—State-of-the-art and the challenge. *Comput. Speech Lang.* **27**, 4–39 (2013).
57. B. Schuller, S. Steidl, A. Batliner, J. Epps, F. Eyben, F. Ringeval, E. Marchi, Y. Zhang, in *15th Annual Conference of the International Speech Communication Association (INTERSPEECH 2014)* (International Speech Communication Association, 2014), pp. 427–431.
58. K. R. Hall, I. DeVore, in *Primate Behavior: Field Studies of Monkeys and Apes*, I. DeVore, Ed. (Holt, Rinehart and Winston, 1965), pp. 53–110.
59. C. Kemp, A. Rey, T. Legou, L.-J. Boë, F. Berthommier, Y. Becker, J. Fagot, in *Origins of Human Language: Continuities and Discontinuities with Nonhuman Primates*, L.-J. Boë, J. Fagot, P. Perrier, J.-L. Schwartz, Eds. (Peter Lang, Frankfurt am Main, 2018), pp. 15–58.
60. J. Fischer, K. Hammerschmidt, D. L. Cheney, R. M. Seyfarth, Acoustic features of female chacma baboon barks. *Ethology* **107**, 33–54 (2001).
61. P. Maciej, I. Ndao, K. Hammerschmidt, J. Fischer, Vocal communication in a complex multi-level society: Constrained acoustic structure and flexible call usage in Guinea baboons. *Front. Zool.* **10**, 58 (2013).
62. M. Briseño-Jaramillo, V. Biquand, A. Estrada, A. Lemasson, Vocal repertoire of free-ranging black howler monkeys' (*Alouatta pigra*): Call types, contexts, and sex-related contributions. *Am. J. Primatol.* **79**, e22630 (2017).
63. M. Benítez, A. Roux, J. Fischer, J. Beehner, T. Bergman, Acoustic and temporal variation in gelada (*Theropithecus gelada*) loud calls advertise male quality. *Int. J. Primatol.* **37**, 568–585 (2016).
64. W. T. Fitch, The evolution of speech: A comparative review. *Trends Cogn. Sci.* **4**, 258–267 (2000).
65. W. T. Fitch, D. Reby, The descended larynx is not uniquely human. *Proc. R. Soc. Lond. B Biol. Sci.* **268**, 1669–1675 (2001).
66. R. Frey, T. Riede, Sexual dimorphism of the larynx of the Mongolian gazelle (*Procapra gutturosa* Pallas, 1777) (Mammalia, Artiodactyla, Bovidae). *Zool. Anz. J. Comp. Zool.* **242**, 33–62 (2003).
67. G. E. Weissengruber, G. Forstenpointner, G. Peters, A. Kübber-Heiss, W. T. Fitch, Hyoid apparatus and pharynx in the lion (*Panthera leo*), jaguar (*Panthera onca*), tiger (*Panthera tigris*), cheetah (*Acinonyx jubatus*) and domestic cat (*Felis silvestris f. catus*). *J. Anat.* **201**, 195–209 (2002).
68. T. Nishimura, A. Mikami, J. Suzuki, T. Matsuzawa, Descent of the larynx in chimpanzee infants. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 6930–6933 (2003).
69. T. Nishimura, Comparative morphology of the hyo-laryngeal complex in anthropoids: Two steps in the evolution of the descent of the larynx. *Primates* **44**, 41–49 (2003).
70. W. T. Fitch, The phonetic potential of nonhuman vocal tracts: Comparative cineradiographic observations of vocalizing animals. *Phonetica* **57**, 205–218 (2000).
71. D. Rendall, S. Kollias, C. Ney, P. Lloyd, Pitch (F0) and formant profiles of human vowels and vowel-like baboon grunts: The role of vocalizer body size and voice-acoustic allometry. *J. Acoust. Soc. Am.* **117**, 944–955 (2005).

72. W. T. Fitch, *The Evolution of Language* (Cambridge Univ. Press, Cambridge, 2010).
73. W. T. Fitch, M. D. Hauser, Vocal production in nonhuman primates: Acoustics, physiology, and functional constraints on "honest" advertisement. *Am. J. Primatol.* **37**, 191–219 (1995).
74. W. T. Fitch, M. D. Hauser, in *Acoustic Communication*, A. M. Simmons, A. N. Popper, R. R. Fay, Eds. (Springer, 2003), pp. 65–137.
75. D. Rendall, Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons. *J. Acoust. Soc. Am.* **113**, 3390–3402 (2003).
76. R. D. Hienz, J. V. Brady, The acquisition of vowel discriminations by nonhuman primates. *J. Acoust. Soc. Am.* **84**, 186–194 (1988).
77. M. S. Sommers, D. B. Moody, C. A. Prosen, W. C. Stebbins, Formant frequency discrimination by Japanese macaques (*Macaca fuscata*). *J. Acoust. Soc. Am.* **91**, 3499–3510 (1992).
78. D. Rendall, thesis, University of California, Davis (1996).
79. J. Fischer, Barbary macaques categorize shrill barks into two call types. *Anim. Behav.* **55**, 799–807 (1998).
80. M. D. Hauser, Functional referents and acoustic similarity: Field playback experiments with rhesus monkeys. *Anim. Behav.* **55**, 1647–1658 (1998).
81. W. T. Fitch, J. B. Fritz, Rhesus macaques spontaneously perceive formants in conspecific vocalizations. *J. Acoust. Soc. Am.* **120**, 2132–2141 (2006).
82. W. T. Fitch, Skull dimensions in relation to body size in nonhuman mammals: The causal bases for acoustic allometry. *Fortschr. Zool.* **103**, 40–58 (2000).
83. H. Hatano, T. Kitamura, H. Takemoto, P. Mokhtari, K. Honda, S. Masaki, in *13th Annual Conference of the International Speech Communication Association (INTERSPEECH 2012)* (International Speech Communication Association, 2012), pp. 402–405.
84. P. Lieberman, *On the Origins of Language: An Introduction to the Evolution of Human Speech* (Macmillan, 1975).
85. M. D. Hauser, C. S. Evans, P. Marler, The role of articulation in the production of rhesus monkey, *Macaca mulatta*, vocalizations. *Anim. Behav.* **45**, 423–433 (1993).
86. P. Rubin, E. Vatikiotis-Bateson, in *Animal Acoustic Communication*, S. L. Hopp, M. J. Owren, C. S. Evans, Eds. (Springer, 1998), pp. 251–290.
87. S. Lee, A. Potamianos, S. Narayanan, Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *J. Acoust. Soc. Am.* **105**, 1455–1468 (1999).
88. T. Riede, E. Bronson, H. Hatzikirou, K. Zuberbühler, Vocal production mechanisms in a non-human primate: Morphological data and a model. *J. Hum. Evol.* **48**, 85–96 (2005).
89. L.-J. Boë, J.-L. Heim, K. Honda, S. Maeda, The potential Neandertal vowel space was as large as that of modern humans. *J. Phon.* **30**, 465–484 (2002).
90. K. Honda, M. K. Tiede, in *The 5th International Conference on Spoken Language Processing (ICSLP-1998)* (International Speech Communication Association, 1998), p. paper 0686.
91. J.-L. Heim, *Les Hommes Fossiles de La Ferrassie/1, Le Gisement. Les Squelettes Adultes (Crâne et Squelette du Tronc)*, vol. 1 of *Archives de l'Institut de Paléontologie Humaine* (Masson, 1976).
92. A. Barney, S. Martelli, A. Serrurier, J. Steele, Articulatory capacity of Neanderthals, a very recent and human-like fossil hominin. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 88–102 (2012).
93. W. T. Fitch, R. A. Suthers, in *Vertebrate Sound Production and Acoustic Communication*, R. A. Suthers, W. T. Fitch, R. R. Fay, A. N. Popper, Eds. (Springer, 2016), pp. 1–18.
94. W. J. Hardcastle, J. Laver, in *The Handbook of Phonetic Sciences* (Blackwell, 1997).
95. P. Ladefoged, K. Johnson, *A Course in Phonetics* (Cengage, ed. 7, 2014).
96. R. K. Potter, G. A. Kopp, H. C. Green, *Visible Speech* (Van Nostrand, 1947).
97. G. Fant, *Acoustic Theory of Speech Production: With Calculations Based on X-Ray Studies of Russian Articulations* (Mouton, 1960).
98. B. H. Story, I. R. Titze, E. A. Hoffman, Vocal tract area functions from magnetic resonance imaging. *J. Acoust. Soc. Am.* **100**, 537–554 (1996).
99. M. M. Sondhi, Resonances of a bent vocal tract. *J. Acoust. Soc. Am.* **79**, 1113–1116 (1986).
100. A. M. Taylor, D. Reby, The contribution of source–filter theory to mammal vocal communication research. *J. Zool.* **280**, 221–236 (2010).
101. A. M. Taylor, B. D. Charlton, D. Reby, *Vertebrate Sound Production and Acoustic Communication*, R. A. Suthers, W. T. Fitch, R. R. Fay, A. N. Popper, Eds. (Springer, 2016), pp. 229–259.
102. D. Pfefferle, J. Fischer, Sounds and size: Identification of acoustic variables that reflect body size in hamadryas baboons, *Papio hamadryas*. *Anim. Behav.* **72**, 43–51 (2006).
103. J. D. Markel, A. H. Gray, in *Linear Prediction of Speech* (Springer-Verlag, 1976).
104. M. J. Owren, R. H. Bernacki, in *Animal Acoustic Communication* (Springer, 1998), pp. 129–162.
105. K. Johnson, *Acoustic and Auditory Phonetics* (Wiley-Blackwell, 2012).
106. B. de Boer, in *Acoustics '08 Paris* (Société Française d'Acoustique, 2008), pp. 8695–8700.
107. B. S. Atal, J. J. Chang, M. V. Mathews, J. W. Tukey, Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique. *J. Acoust. Soc. Am.* **63**, 1535–1555 (1978).
108. L.-J. Boë, P. Perrier, G. Bailly, The geometric vocal tract variables controlled for vowel production: Proposals for constraining acoustic-to-articulatory inversion. *J. Phon.* **20**, 27–38 (1992).
109. B. Lindblom, J. Sundberg, Acoustical consequences of lip, tongue, jaw, and larynx movement. *J. Acoust. Soc. Am.* **50**, 1166–1179 (1971).
110. R. Harshman, P. Ladefoged, L. Goldstein, Factor analysis of tongue shapes. *J. Acoust. Soc. Am.* **62**, 693–707 (1977).
111. S. Maeda, in *Speech Production & Speech Modelling*, W. J. Hardcastle, A. Marchal, Eds. (Kluwer, Dordrecht, 1990), pp. 131–149.
112. A. Bothorel, P. Simon, F. Wioland, J.-P. Zerling, *Cinéradiographie des Voyelles et Consonnes du Français* (Institut de Phonétique de Strasbourg, 1986).
113. J. A. V. Valdés Vargas, P. Badin, L. Lamalle, in *13th Annual Conference of the International Speech Communication Association (INTERSPEECH 2012)* (International Speech Communication Association, 2012), pp. 2186–2189.
114. S. Maeda, K. Honda, From EMG to formant patterns of vowels: The implication of vowel spaces. *Phonetica* **51**, 17–29 (1994).
115. K. Honda, Organization of tongue articulation for vowels. *J. Phon.* **24**, 39–52 (1996).
116. S. Waltl, P. Hoole, in *8th International Seminar on Speech Production*, R. Sock, S. Fuchs, Y. Laprie, Eds. (INRIA, 2008), pp. 445–448.
117. S. Buchaillard, P. Perrier, Y. Payan, A biomechanical model of cardinal vowel production: Muscle activations and the impact of gravity on tongue positioning. *J. Acoust. Soc. Am.* **126**, 2033–2051 (2009).
118. B. Lindblom, J. Sundberg, A quantitative model of vowel production and the distinctive features of Swedish vowels. *Speech Transm. Lab. Q. Prog. Status Rep.* **10**, 14–32 (1969).
119. L.-J. Boë, P. Perrier, B. Guérin, J.-L. Schwartz, in *First European Conference on Speech Communication and Technology* (International Speech Communication Association, 1989), pp. 2281–2284.
120. P. Mermelstein, Articulatory model for the study of speech production. *J. Acoust. Soc. Am.* **53**, 1070–1082 (1973).
121. L.-J. Boë, S. Maeda, in *Journées d'Études Linguistique* (Université de Nantes, 1997), pp. 98–105.
122. L. Ménard, thesis, Université Stendhal, Grenoble III, Grenoble, France (2002).
123. G. Barbier, L.-J. Boë, G. Captier, La croissance du conduit vocal du foetus à l'adulte: Une étude longitudinale. *Biométrie Hum. Anthropol.* **30**, 11–22 (2011).
124. G. Barbier, L.-J. Boë, G. Captier, R. Laboissière, in *16th Annual Conference of the International Speech Communication Association (INTERSPEECH 2015)* (International Speech Communication Association, 2015), pp. 364–368.
125. H. Mol, *Fundamentals of Phonetics: II, Acoustical Models Generating the Formants of the Vowel Phonemes* (Mouton, 1970).
126. R. D. Kent, A. D. Murray, Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *J. Acoust. Soc. Am.* **72**, 353–365 (1982).
127. P. Lieberman, *The Biology and Evolution of Language* (Harvard Univ. Press, 1984).
128. B. de Boysson-Bardies, P. Halle, L. Sagart, C. Durand, A crosslinguistic investigation of vowel formants in babbling. *J. Child Lang.* **16**, 1–17 (1989).
129. C. L. Matyear, P. F. MacNeilage, B. L. Davis, Nasalization of vowels in nasal environments in babbling: Evidence for frame dominance. *Phonetica* **55**, 1–17 (1998).
130. L. Ménard, L.-J. Boë, L'émergence du système phonologique chez l'enfant: L'apport de la modélisation articulatoire. *Can. J. Linguist. Rev. Can. Linguist.* **49**, 155–174 (2004).
131. L. Ménard, J.-L. Schwartz, L.-J. Boë, Role of vocal tract morphology in speech development: Perceptual targets and sensorimotor maps for synthesized French vowels from birth to adulthood. *J. Speech Lang. Hear. Res.* **47**, 1059–1080 (2004).
132. L. Ménard, J.-L. Schwartz, L.-J. Boë, J. Aubin, Articulatory–acoustic relationships during vocal tract growth for French vowels: Analysis of real data and simulations with an articulatory model. *J. Phon.* **35**, 1–19 (2007).
133. H. K. Vorperian, R. D. Kent, Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *J. Speech Lang. Hear. Res.* **50**, 1510–1545 (2007).
134. L. Ménard, B. L. Davis, L.-J. Boë, J.-P. Roy, Producing American English vowels during vocal tract growth: A perceptual categorization study of synthesized vowels. *J. Speech Lang. Hear. Res.* **52**, 1268–1285 (2009).
135. H. K. Vorperian, S. Wang, M. K. Chung, E. M. Schimek, R. B. Durtschi, R. D. Kent, A. J. Ziegert, L. R. Gentry, Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study. *J. Acoust. Soc. Am.* **125**, 1666–1678 (2009).
136. H. K. Vorperian, S. Wang, E. M. Schimek, R. B. Durtschi, R. D. Kent, L. R. Gentry, M. K. Chung, Developmental sexual dimorphism of the oral and pharyngeal portions of the vocal tract: An imaging study. *J. Speech Lang. Hear. Res.* **54**, 995–1010 (2011).
137. P. K. Kuhl, A. N. Meltzoff, Infant vocalizations in response to speech: Vocal imitation and developmental change. *J. Acoust. Soc. Am.* **100**, 2425–2438 (1996).
138. B. de Boer, Modelling vocal anatomy's significant effect on speech. *J. Evol. Psychol.* **8**, 351–366 (2010).
139. P. Badin, L.-J. Boë, T. R. Sawallis, J.-L. Schwartz, Keep the lips to free the larynx: Comments on de Boer's articulatory model (2010). *J. Phon.* **46**, 161–167 (2014).
140. B. H. Story, Comparison of magnetic resonance imaging-based vocal tract area functions obtained from the same speaker in 1994 and 2002. *J. Acoust. Soc. Am.* **123**, 327–335 (2008).

141. D. Reby, K. McComb, Anatomical constraints generate honesty: Acoustic cues to age and weight in the roars of red deer stags. *Anim. Behav.* **65**, 519–530 (2003).
142. W. T. Fitch, B. de Boer, N. Mathur, A. A. Ghazanfar, Monkey vocal tracts are speech-ready. *Sci. Adv.* **2**, e1600723 (2016).
143. M. Kelly, *Monkey Speak: Macaques Have the Anatomy, Not the Brain, for Human Speech* (Princeton University News, 2016); www.princeton.edu/news/2016/12/09/monkey-speak-macaques-have-anatomy-not-brain-human-speech).
144. B. de Boer, W. T. Fitch, Computer models of vocal tract evolution: An overview and critique. *Adapt. Behav.* **18**, 36–47 (2010).
145. B. de Boer, Investigating the acoustic effect of the descended larynx with articulatory models. *J. Phon.* **38**, 679–686 (2010).
146. J. Fagot, L.-J. Boë, F. Berthommier, N. Claidière, R. Malassis, A. Meguerditchian, A. Rey, M. Montant, The baboon: A model for the study of language evolution. *J. Hum. Evol.* **126**, 39–50 (2019).
147. R. D'Anastasio, S. Wroe, C. Tuniz, L. Mancini, D. T. Cesana, D. Dreossi, M. Ravichandiran, M. Attard, W. C. H. Parr, A. Agur, L. Capasso, Micro-biomechanics of the Kebara 2 hyoid and its implications for speech in Neanderthals. *PLOS ONE* **8**, e82261 (2013).
148. J.-J. Hublin, A. Ben-Ncer, S. E. Bailey, S. E. Freidline, S. Neubauer, M. M. Skinner, I. Bergmann, A. Le Cabec, S. Benazzi, K. Harvati, P. Gunz, New fossils from Jebel Irhoud, Morocco and the pan-African origin of *Homo sapiens*. *Nature* **546**, 289–292 (2017).
149. D. E. Lieberman, B. M. McBratney, G. Krovitz, The evolution and development of cranial form in *Homo sapiens*. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 1134–1139 (2002).
150. D. Gokhman, L. Agranat-Tamir, G. Housman, R. García-Pérez, M. Nissim-Rafinia, S. Mallick, M. A. Nieves-Colón, H. Li, S. Alpaslan-Roodenberg, M. Novak, H. Gu, M. Ferrando-Bernal, P. Gelabert, I. Lipende, I. Kondova, R. Bontrop, E. E. Quillen, A. Meissner, A. C. Stone, A. E. Pusey, D. Mjungu, L. Kandel, M. Liebergall, M. E. Prada, J. M. Vidal, K. Prüfer, J. Krause, B. Yakir, S. Pääbo, R. Pinhasi, C. Lalueza-Fox, D. Reich, T. Marques-Bonet, E. Meshorer, L. Carmel, Extensive regulatory changes in genes affecting vocal and facial anatomy separate modern from archaic humans. bioRxiv 106955 [**Preprint**] 3 October 2017.
151. R. Quam, I. Martínez, M. Rosa, A. Bonmatí, C. Lorenzo, D. J. de Ruiter, J. Moggi-Cecchi, M. Conde Valverde, P. Jarabo, C. G. Menter, J. F. Thackeray, J. L. Arsuaga, Early hominin auditory capacities. *Sci. Adv.* **1**, e1500355 (2015).
152. J. B. Palmer, N. J. Rudin, G. Lara, A. W. Crompton, Coordination of mastication and swallowing. *Dysphagia* **7**, 187–200 (1992).
153. G. Hickok, A cortical circuit for voluntary laryngeal control: Implications for the evolution language. *Psychon. Bull. Rev.* **24**, 56–63 (2017).
154. V. Kumar, K. Simonyan, in *Origins of Human Language: Continuities and Discontinuities with Nonhuman Primates*, L.-J. Boë, J. Fagot, P. Perrier, J.-L. Schwartz, Eds. (Peter Lang, 2018), pp. 137–151.
155. A. D. Friederici, Evolution of the neural language network. *Psychon. Bull. Rev.* **24**, 41–47 (2017).
156. W. D. Hopkins, in *Origins of Human Language: Continuities and Discontinuities with Nonhuman Primates*, L.-J. Boë, J. Fagot, P. Perrier, J.-L. Schwartz, Eds. (Peter Lang, 2018), pp. 153–186.
157. L.-J. Boë, J. Fagot, P. Perrier, J.-L. Schwartz, Eds., *Origins of Human Language: Continuities and Discontinuities with Nonhuman Primates* (Peter Lang, 2018).
158. W. T. Fitch, Preface to the special issue on the biology and evolution of language. *Psychon. Bull. Rev.* **24**, 1–2 (2017).
159. M. A. Arbib, From mirror neurons to complex imitation in the evolution of language and tool use. *Ann. Rev. Anthropol.* **40**, 257–273 (2011).
160. P. F. MacNeilage, The frame/content theory of evolution of speech production. *Behav. Brain Sci.* **21**, 499–546 (1998).
161. P. F. MacNeilage, B. L. Davis, Motor mechanisms in speech ontogeny: Phylogenetic, neurobiological and linguistic implications. *Curr. Opin. Neurobiol.* **11**, 696–700 (2001).
162. M. D. Hauser, C. Yang, R. C. Berwick, I. Tattersall, M. J. Ryan, J. Watumull, N. Chomsky, R. C. Lewontin, The mystery of language evolution. *Front. Psychol.* **5**, 401 (2014).
163. S. C. Levinson, Turn-taking in human communication: Origins and implications for language processing. *Trends Cogn. Sci.* **20**, 6–14 (2016).
164. C. T. Snowdon, Learning from monkey "talk". *Science* **355**, 1120–1122 (2017).
165. A. R. Lameira, M. E. Hardus, A. Mielke, S. A. Wich, R. W. Shumaker, Vocal fold control beyond the species-specific repertoire in an orang-utan. *Sci. Rep.* **6**, 30315 (2016).
166. R. M. Seyfarth, D. L. Cheney, Animal cognition: Chimpanzee alarm calls depend on what others know. *Curr. Biol.* **22**, R51–R52 (2012).
167. C. Crockford, R. M. Wittig, K. Zuberbühler, Vocalizing in chimpanzees is influenced by social-cognitive processes. *Sci. Adv.* **3**, e1701742 (2017).
168. J. P. Taglialatela, L. Reamer, S. J. Schapiro, W. D. Hopkins, Social learning of a communicative signal in captive chimpanzees. *Biol. Lett.* **8**, 498–501 (2012).
169. N. Claidière, K. Smith, S. Kirby, J. Fagot, Cultural evolution of systematically structured behaviour in a non-human primate. *Proc. Biol. Sci.* **281**, 20141541 (2014).
170. C. Saldana, J. Fagot, S. Kirby, K. Smith, N. Claidière, High-fidelity copying is not necessarily the key to cumulative cultural evolution: A study in monkeys and children. *Proc. Biol. Sci.* **286**, 20190729 (2019).
171. P. Boersma, D. Weenink, *Praat: Doing phonetics by computer* (2018); www.praat.org.
172. E. Zee, in *15th International Congress of Phonetic Sciences (ICPhS-15)*. ICPhS Archives at the IPA (www.internationalphoneticassociation.org) (2003), pp. 1117–1120.
173. L. C. W. Pols, H. R. C. Tromp, R. Plomp, Frequency analysis of Dutch vowels from 50 male speakers. *J. Acoust. Soc. Am.* **53**, 1093–1101 (1973).
174. J. Hillenbrand, L. A. Getty, M. J. Clark, K. Wheeler, Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am.* **97**, 3099–3111 (1995).
175. Calliope, *La Parole et Son Traitement Automatique* (Masson, 1989).
176. C. Meunier, in *Les Dysarthries, P. Auzou*, Ed. (Solal, 2007), pp. 164–173.
177. M. Pätzold, A. P. Simpson, in *The Kiel Corpus of Read/Spontaneous Speech: Acoustic Data Base, Processing Tools and Analysis Results*, A. P. Simpson, K. J. Kohler, T. Rettstadt, Eds. (IPDS, Universität Kiel, 1996), pp. 215–247.
178. C.-S. Yang, H. Kasuya, in *The 4th International Conference on Spoken Language Processing. ICSLP '96* (IEEE, 1996), vol. 2, pp. 949–952.
179. H. N. Ting, C. H. Lee, in *IFMBE Proceedings* (Springer, 2012), pp. 375–378.
180. P. Escudero, P. Boersma, A. S. Rauber, R. A. H. Bion, A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *J. Acoust. Soc. Am.* **126**, 1379–1393 (2009).
181. M. Contini, L. J. Boë, Voyelles orales et nasales du sarde campidanien: Étude acoustique et phonologique. *Phonetica* **25**, 165–191 (1972).
182. G. Fant, Acoustic analysis and synthesis of speech with applications to Swedish. *Ericsson Tech.* **15**, 3–108 (1959).
183. G. Fant, G. Henningsson, U. Stålhammar, Formant frequencies of Swedish vowels. *Speech Transm. Lab. Q. Prog. Status Rep.* **10**, 26–31 (1969).
184. H. Pineau, La croissance et ses lois, Laboratoire d'Anatomie de la Faculté de Médicine, Paris (1965).
185. P. Badin, G. Fant, Notes on vocal tract computation. *Speech Transm. Lab. Q. Prog. Status Rep.* **25**, 53–108 (1984).
186. B. Nadhurou, M. Gamba, N. V. Andriaholinirina, A. Ouledi, C. Giacoma, The vocal communication of the mongoose lemur (*Eulemur mongoz*): Phonation mechanisms, acoustic features and quantitative analysis. *Ethol. Ecol. Evol.* **28**, 241–260 (2016).
187. D. Rendall, M. J. Owren, E. Weerts, R. D. Hienz, Sex differences in the acoustic structure of vowel-like grunt vocalizations in baboons and their perceptual discrimination by baboon listeners. *J. Acoust. Soc. Am.* **115**, 411–421 (2004).

Acknowledgments

Funding: This work was supported by the European Research Council under the Seventh European Community Program [FP7/2007–2013 grant agreement no. 339152—"Speech Unit(e)s"] and grants ANR-16-CONV-0002 (ILCB), ANR-11-LABX-0036 (BLRI), and the Excellence Initiative of Aix-Marseille University (A*MIDEX). **Author contributions:** L.-J.B. and T.R.S. conceived and wrote the paper, in regular consultation with J.-L.S. and J.F. who also contributed passages; L.-J.B. developed the figures with assistance from T.R.S. and P.B.; all authors provided materials used in the final paper, particularly P.B. for acoustics, L.M. and G.B. for human speech development, J.-L.H. for Neanderthal anatomy, G.C. for primate anatomy, and J.F. for primates in general; all living authors approved the final text. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper. Questions about additional materials related to this paper may be directed to the authors.

Submitted 16 December 2018
Accepted 10 October 2019
Published 11 December 2019
10.1126/sciadv.aaw3916

Citation: L.-J. Boë, T. R. Sawallis, J. Fagot, P. Badin, G. Barbier, G. Captier, L. Ménard, J.-L. Heim, J.-L. Schwartz, Which way to the dawn of speech?: Reanalyzing half a century of debates and data in light of speech science. *Sci. Adv.* **5**, eaaw3916 (2019).