Review of the Origin and Evolution of Human Speech: Postprint

Authors: Kong Jiangping, Wu Xiyu

Date: 2021-01-18T00:00:00+00:00

Abstract

This article comprehensively reviews the research achievements on the origin and evolution of human speech, provides scholars and graduate students interested in the origin and evolution of speech with a substantial body of relevant research literature, and offers a preliminary commentary on previous studies in this field.

Full Text

Preamble

A Review on the Origin and Evolution of Human Speech
Kong Jiangping, Wu Xiyu

Abstract

This paper reviews the research results on the origin and evolution of human speech in detail, provides a large number of relevant literature references for scholars and graduate students interested in the origin and evolution of speech, and offers a preliminary review of previous studies in this field.

Keywords: the origin of speech; the evolution of speech; the theory of laryngeal descent; the vocalization of primates; fossil hominins

Language plays an important role in human evolution, and finding evidence to reveal the origins of language has long been a goal pursued by paleoanthropologists, mammalogists, phoneticians, linguists, and speech scientists. The study of language origins is a multidisciplinary science that integrates paleoanthropology, anatomy, phonetics, speech science, cognitive science, and related fields. In research on language evolution, the study of speech production is the most feasible because it employs empirical scientific methods. Currently, the dominant framework is Lieberman' s laryngeal descent theory (LDT). Lieberman first studied the evolution of speech from a biological perspective, and later examined the evolutionary process from multiple angles: human speech, the vocal abilities of existing non-human primates, fossil records of human speech evolution, the supralaryngeal vocal tract (SVT), and the neural mechanisms of speech, ultimately proposing the laryngeal descent theory ([1]Lieberman,1984; [2]Lieberman et al,2015). Lieberman' s laryngeal descent theory has dominated the field for nearly half a century.

However, in the past decade, the laryngeal descent theory has been challenged due to the accumulation of new evidence and the emergence of new perspectives ([3]Fitch, 2000). These new evidence and debates have focused on four main questions: First, can laryngeal descent in mammals and non-human primates produce human speech sounds ([4]Fitch et al, 2001)? Second, are the vocal tracts of extant non-human primates uniformly shaped, and can they produce sounds with contrasting formant patterns without laryngeal descent ([5]Boë et al, 2017)? Third, regarding methodology, is it effective to use vocal tract models to simulate vowel evolution and emergence? Fourth, how should vowel emergence be defined and when did it occur? Boë and colleagues have provided excellent commentary on these debates, new perspectives, and research methods ([6]Boë et al, 2019). Below, we provide a review and synthesis of research on the origin and evolution of human speech.

2. Laryngeal Descent Theory

In the 1950s, Lieberman began using new techniques and concepts to study vocalizations in non-human primates, establishing himself as one of the pioneers in this research area. Starting in the 1970s, Lieberman and colleagues used spectral analysis to examine the calls of macaques, chimpanzees, and gorillas, investigating the relationship between vocal tract anatomy and acoustic characteristics based on these analyses. They concluded that these animals' vocalizations were primarily similar to central vowel sounds, suggesting that non-human primates lacked the ability to change the shape of the supralaryngeal vocal tract during vocalization ([7]Lieberman, 1968). Because macaques and apes have very small pharyngeal cavities, they cannot produce human speech sounds due to fundamental physiological constraints, making it impossible to generate the vocal tract area functions required for continuously variable speech ([8]Lieberman et al, 1969).

Lieberman also studied the cries of newborn human infants and found that, like other non-human primates, human infants do not use the supralaryngeal vocal tract during crying except for overall laryngeal movement. Due to limitations of the vocal organs, they also lack variation in pharyngeal cavity area ([9]Lieberman et al, 1971). Lieberman and Crelin further used fossil models to infer the anatomical structure and shape of the vocal tract in Neanderthals, estimating their speech production potential and comparing it with that of modern humans ([10]Lieberman et al, 1971). Lieberman' s research demonstrated that the function of the vocal tract extends beyond breathing and swallowing, as skeletal evidence from human evolution reveals a series of changes that likely adapted the primate vocal tract for speech production to some degree. Therefore, these changes should be considered part of the human speech output mechanism —specifically, a uniquely human vocal apparatus.

Another important study on vocal tract evolution approached the issue from the perspective of fossil hominins. The divergence between chimpanzees and modern humans can be traced back to approximately 6 million years ago ([11]Vallender Eric, et al. 2008). However, obtaining fossil materials is extremely difficult, which has constrained research on vocal tract evolution from ancient human fossils. Based on the major fossil hominins available worldwide, only about a dozen specimens preserve both cranium (skull), upper mandible, upper teeth, lower mandible, and lower teeth together: Pataud 1, Combe-Capelle, Dolni Vestonice XV and XVI, Barma Grande 5, La Ferrassie 1, Le Moustier, Pech De L' aze, Roc de Marsal, Fish Hoek, Nariokotome, Amud, Jebei Qafzeh 9, Jebei Qafzeh 11, Shanidar 1, Skhul V, Tabun I, Teshik-Tash, and Zhoukoudian Upper Cave: Cast of PA.101. In contrast, relatively more complete fossil crania are available, numbering several dozen ([12]Jeffrey et al, 2002; [13]Jeffrey et al, 2003). Since muscles in the oral cavity and laryngeal cartilages do not fossilize, this creates significant challenges for reconstructing the vocal tract.

In studies using fossils to investigate human vocal tract evolution, researchers first measured and compared the developmental process of laryngeal descent in modern human infants and adults ([14]Laitman et al, 1976). They simultaneously measured and studied several key points on the cranial base of primate and hominin fossils: the alveolar midpoint (A, prosthion), posterior oral point (B, staphylion), spheno-occipital point (C, hormion), sphenoid point (D, sphenobasion), and cranial base point (E, endobasion). The research found that the suprapharyngeal constrictor muscles run forward and downward almost perpendicular to the defined line DE on the occipital plane. The study also compared distances between these six points on the crania of hominin and non-hominin primates, discussing the relationship between these distances and laryngeal descent ([15]Laitman et al, 1978; [16]Laitman et al, 1979; [17]Laitman et al, 1982). Based on these fossil studies, Budil developed a method for calculating these parameters, proposed the K parameter, and computed K values for the crania of australopithecines, Neanderthals, and modern humans. On this basis, he performed functional reconstructions of the vocal tract in hominins ([18]Budil, 1994).

The research by Lieberman and colleagues on the physiological morphology and developmental processes of the vocal organs in chimpanzees, modern human infants, and adults, along with Laitman' s work on hominin fossils, both provided evidence for Lieberman' s laryngeal descent theory (LDT). Within this theoretical framework, numerous studies have investigated vocal tract evolution from chimpanzees to modern humans, reconstructing ancient human vocal tracts and concluding that laryngeal descent represents the most important evolutionary pathway for human vocal tract evolution and speech production ([1]Lieberman, 1984; [2]Lieberman et al, 2015). Based on solid foundations of acoustic and anatomical observation, Lieberman' s conclusions were considered factual and served as the touchstone for this research field, eventually becoming known as the laryngeal descent theory (LDT).

3. Laryngeal Descent in Non-Human Primates and Mammals

Research on mammalian larynges has revealed that laryngeal descent also occurs in some mammals, including deer ([4]Fitch et al, 2001), Mongolian gazelles ([19]Frey et al, 2003), lions, jaguars, tigers, cheetahs, and domestic cats ([20]Weissengruber et al, 2002), chimpanzees ([21]Nishimura et al, 2003; [22]Nishimura, 2005), and other primates ([23]Nishimura, 2003). Studies have also examined the continuity and discontinuity of vocalizations in non-human primates ([24]Berthommier et al, 2018). Fitch documented and recorded the dynamic and temporary processes of laryngeal descent, hyoid bone movement, and tongue positioning in goats, dogs, pigs, and marmosets, arguing that laryngeal descent is not unique to humans and suggesting two distinct evolutionary pathways for this phenomenon ([3]Fitch, 2000). He proposed that laryngeal descent serves to lengthen the vocal tract, lowering its resonant frequencies to create the perception of exaggerated body size, thereby providing a non-linguistic alternative hypothesis ([25]Fitch, 1997). Fitch argued that, like all mammals, early hominins could reconfigure their vocal anatomy during vocalization to obtain a vocal tract structure sufficient for producing any intelligible phoneme. However, the vocal tract morphology resulting from laryngeal descent in non-human animals still differs substantially from that of humans.

4. Vocalizations of Non-Human Primates

It is commonly believed that only humans possess vocal organs capable of linguistic articulation, and it has been assumed that the vocal tracts of mammals, including non-human primates, resemble a uniform or open tube during vocalization ([7]Lieberman, 1968; [8]Lieberman et al, 1969; [26]Shipley et al, 1991). However, recent research suggests that non-human primates may also be capable of producing different vowels. Studies on baboon vocalizations indicate that baboon vocal tracts are not uniform ([27]Owren et al, 1997). Riede and colleagues first tested the consistency between Diana monkey alarm calls and tube characteristics, finding that their eagle and leopard alarm calls did not resemble schwa but were similar to human vowel /a/ produced by 10- to 12-year-old children ([28]Riede et al, 2005). These results correlate with human vowels produced by children with similar vocal tract length (VTL) ([29]Lee, 1999). Fitch and colleagues studied the potential vowels that macaques could produce, concluding that macaque vocal tracts can easily generate a sufficient phonetic range to support spoken language and that they are capable of producing human speech sounds ([30]Fitch et al, 2016). Boë and colleagues' research on baboons (Papio papio) demonstrated that baboons can naturally produce vowels with formant structures similar to those of humans ([5]Boë et al, 2017). Pisanski and colleagues suggested this might represent a living remnant of early vocal control abilities that led to articulate human speech ([31]Pisanski et al, 2016). Fitch attributed macaques' inability to speak to their lack of evolved neural systems capable of producing language, arguing that the evolution of human language capacity requires changes in the nervous system rather than modifications to vocal anatomy ([30]Fitch et al, 2016). Based on research on vocal tract anatomy and articulation in non-human primates, some scholars support the "neural hypothesis," proposing that non-human primates have no problem producing vowels anatomically but lack the corresponding neural mechanisms ([32]Boer, 2010; [33]Boer et al, 2010; [34]Kelly, 2016). However, Lieberman argued that the speech studied by Fitch and colleagues represents macaque vocalizations, which cannot be considered fully human speech in the complete sense ([35]Lieberman, 2017).

5. Vocal Tract Modeling Research

Some scholars argue that biologists have encountered difficulties when attempting to apply principles and methods from language studies to animal communication ([30]Fitch et al, 2016). Based on new results from different vocal tract models and speech production research, our understanding of the relationship between the vocal tract and speech production has deepened. Building on the work of Lindblom and Harshman ([36]Lindblom et al, 1971; [37]Harshman et al, 1977), Maeda' s model improved the sagittal variables of the vocal tract using principal component analysis, with seven control parameters explaining 70%-90% of vowel variation ([38]Maeda, 1990). Using this model and electromyographic signals, researchers have studied the relationship between vocal tract model control parameters and tongue muscle activity ([39]Maeda et al, 1994; [40]Honda, 1996) as well as physiological mechanisms ([41]Buchaillard et al, 2009). Goldstein integrated extensive anatomical data into a clear vocal tract growth model ([42]Goldstein, 1980), employing three parameters: horizontal vocal tract length, vertical vocal tract length, and midline vocal tract length. Goldstein developed a sagittal vocal tract growth model based on her dataset and Mermelstein' s adult model ([43]Mermelstein, 1973), providing area functions and simulated formant values to study vocalizations in newborns, concluding that infants can produce vowels /i/, /a/, and /u/. Boë' s laboratory developed the Variable Linear Articulatory Model (VLAM) to simulate vocal tract growth from birth to adulthood ([44]Boë et al, 1997), introducing two scaling factors for the anterior oral cavity and posterior pharyngeal cavity, with interpolation between these regions. Research shows that even when infants' and children' s pharyngeal cavities are less than half the length of adults, they can begin to establish vowel contrasts before laryngeal descent occurs. Badin' s research demonstrates that lip movements can compensate for the effects of laryngeal descent, enabling normal F1-F2 space even with a high larynx ([45]Badin et al, 2014). Boë and colleagues argue that regardless of pharyngeal cavity length, the size and shape of the pharyngeal tube remain relatively stable from birth to adulthood ([46]Boë et al, 2013). However, Boer contends that these models are overly simplified ([32]Boer, 2010). Overall, research on vocal tract models shows that continuous improvement of these models can advance simulation studies of vocal tract evolution.

6. Inferences on the Timing of Vowel Emergence

Vowel emergence involves two main aspects: the manner and process of vowel emergence, and the timing of vowel emergence. Current research on the conditions and processes of vowel emergence primarily uses information from ancient human fossils to construct articulatory physiological morphology and models for discussion. Another approach examines vowel emergence through studies of vocalizations in humans and non-human primates, which we have reviewed in previous sections. Regarding the timing of vowel emergence, several viewpoints exist: First, Lieberman, Laitman, Crelin, and colleagues conducted extensive fossil research, concluding that Neanderthals lacked modern human articulatory capabilities and possessed only limited speech abilities based on their vocal morphology. Some scholars argue that modern humans and Neanderthals do not belong to the same lineage and therefore cannot be directly compared. Lieberman and colleagues propose that speech emerged approximately in the late stage of *Homo erectus* (0.1 mya) ([1]Lieberman, 1984; [2]Lieberman et al, 2015; [47]Laitman, 1985; [48]Crelin, 1989). Second, Budil developed a K coefficient based on Laitman and colleagues' measurements of ancient human cranial base fossils and performed reconstructions of ancient human articulatory morphology, speculating that human speech capabilities formed during the Upper Pleistocene (0.126-0.0117 mya) ([18]Budil, 1994). Third, Arensburg studied a complete hyoid bone fossil and concluded that Neanderthal hyoid morphology and articulatory structure showed no difference from modern humans, speculating that speech emerged during the Middle Paleolithic (approximately 0.2 to 0.045 mya) and that the evolution of vocal organs began even earlier, before the Middle Paleolithic ([49]Arensburg et al, 1989). Fourth, Boë and colleagues reviewed current research on speech evolution and origins, arguing that existing evidence overwhelmingly refutes the long-standing laryngeal descent theory and pushes back "the origin of language" from approximately 0.2 mya to about 20 mya ([6]Boë et al, 2019).

7. Conclusion

The research reviewed above demonstrates that both the laryngeal descent theory and the challenges to it with new evidence center on several key issues: First, does laryngeal descent necessarily produce the three basic vowels required for language? Second, can the three basic vowels of language be produced without laryngeal descent? Third, how should vowel formant contrast patterns be determined? Fourth, how can we infer the timing of vowel emergence? Consequently, the laryngeal descent theory and the new research questioning and debating this theory constitute the current state of research on the evolution of the human speech vocal tract.

Kong Jiangping
Department of Chinese Language and Literature, Peking University, Ph.D., Professor. Main research areas: phonetics, speech production, and the origin of speech.
Email: jpkong@pku.edu.cn

Wu Xiyu
Department of Chinese Language and Literature, Peking University, Ph.D., Assistant Professor, Associate Researcher. Main research areas: speech production, physiological models of speech, and speech disorders.
Email: xiyuwu@pku.edu.cn

Additional References

Duchin, L. E. (1990). The evolution of articulate speech: Comparative anatomy of the oral cavity in Pan and Homo. Journal of Human Evolution, 19, 687–697.

H. K. Vorperian, S. Wang, E. M. Schimek, R. B. Durtschi, R. D. Kent, L. R. Gentry, M. K. Chung, 2011, Developmental sexual dimorphism of the oral and pharyngeal portions of the vocal tract: An imaging study. J. Speech Lang. Hear. Res. 54, 995–1010.

H. K. Vorperian, S. Wang, M. K. Chung, E. M. Schimek, R. B. Durtschi, R. D. Kent, A. J. Ziegert, L. R. Gentry, 2009, Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study. J. Acoust. Soc. Am. 125, 1666–1678 (2009).

L. Ménard, B. L. Davis, L.-J. Boë, J.-P. Roy, 2009, Producing American English vowels during vocal tract growth: A perceptual categorization study of synthesized vowels. J. Speech Lang. Hear. Res. 52, 1268–1285 (2009).

L. Ménard, J.-L. Schwartz, L.-J. Boë, 2004, Role of vocal tract morphology in speech development: Perceptual targets and sensorimotor maps synthesized French vowels from birth to adulthood. J. Speech Lang. Hear. Res. 47, 1059-1080 (2004).

L. Ménard, J.-L. Schwartz, L.-J. Boë, J. Aubin, 2007, 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).

R. D. Kent, A. D. Murray, 1982, Acoustic features of infant vocalic utterances at 3, 6, and 9 months. J. Acoust. Soc. Am. 72, 353–365 (1982).

Takemoto H. (2008) Morphological analyses and 3D modeling of tongue musculature of Chimpanzee (Pan troglodytes). Am J Primatol 70:966–975.

S. Adachi, M. Yamada, 1999, An acoustical study of sound production in biphonic singing, Xoömij, J. Acoust. Soc. Am. 105 (5).

Dennis H. Klatt, 1980, Software for a cascade/parallel formant synthesizer, J. Acoust. Soc. Am. 6/(3). Mar. 1980.

Alvin M. Liberman, Katherine Safford Harris, Howard S. Hoffman, and Belver C. Griffith, 1957, The discrimination of speech sounds within and across phoneme boundaries, Journal of Experimental Psychology, Vol. 54, No. 5, 1957.

Liberman, A. M. 1957. Some Results of Research on Speech Perception. The Journal of the Acoustical Society of America, 29(1), 117-123.

B. de Boer, 2010, Investigating the acoustic effect of the descended larynx with articulatory models. J. Phon. 38, 679–686.

P. Lieberman, The evolution of human speech: Its anatomical and neural bases. Curr. Anthropol. 48, 39-66 (2007).

P. Lieberman, The evolution of language and thought. J. Anthropol. Sci. 94, 127-146, (2016).

W. T. Fitch, R. A. Suthers, 2016, In Vertebrate Sound Production and Acoustic Communication, R. A. Suthers, W. T. Fitch, R. R. Fay, A. N. Popper, Eds. (Springer,), pp. 1–18.

B. Lindblom, J. Sundberg, 1969, A quantitative model of vowel production and the distinctive features of Swedish vowels. Speech Transm. Lab. Q. Prog. Status Rep. 10, 14–32.

Crelin, E.S., 1987, The Human Vocal Tract, Vantage New York 1987.

Ian Tattersall, 2008, An Evolutionary Framework for the Acquisition of Symbolic Cognition by Homo sapiens, Comparative Cognition & Behavior Reviews. Vol. 3, pp 99-114.

T. Nishimura, 2003, 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).

D. Rendall, S. Kollias, C. Ney, P. Lloyd, 2005, 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.

Note: Figure translations are in progress. See original paper for figures.