

Evolution of speech and evolution of language

Bart de Boer¹

Published online: 3 August 2016
© Psychonomic Society, Inc. 2016

Abstract Speech is the physical signal used to convey spoken language. Because of its physical nature, speech is both easier to compare with other species' behaviors and easier to study in the fossil record than other aspects of language. Here I argue that convergent fossil evidence indicates adaptations for complex vocalizations at least as early as the common ancestor of Neanderthals and modern humans. Furthermore, I argue that it is unlikely that language evolved separately from speech, but rather that gesture, speech, and song coevolved to provide both a multimodal communication system and a musical system. Moreover, coevolution must also have played a role by allowing both cognitive and anatomical adaptations to language and speech to evolve in parallel. Although such a coevolutionary scenario is complex, it is entirely plausible from a biological point of view.

Keywords Evolution of speech · Speech production · Evolution of language

Speech is the most physical aspect of language, and as such is the most promising aspect to study in the context of the (biological) evolution of language. However, even for speech, individual facts tend to be equivocal, such that there is no consensus about the interpretation of the evidence that exists. Moreover, some researchers consider the physical instantiation of language to be unimportant for the study of its evolution, because the physical dynamics are seen as a separate, ancillary process that is only used for externalization (Chomsky, 2007).

Perhaps it is useful to start by making a case that the vocal tract did indeed undergo selection related to vocal communication. First of all, the human vocal tract is different from that of other primates: not just is the larynx lower, but humans also have a much bigger gap between the larynx and the velum than do other primates, and even than do other mammals with permanently lowered larynges (increasing the risk of choking on one's food; e.g., Heimlich, 1975). Moreover, the human vocal tract may be optimized for vocalization (de Boer, 2010, 2012; but see Badin, Boë, Sawallis, & Schwartz, 2014). Furthermore, as compared to other apes, we lack air sacs (Fitch, 2000) and have better breathing control (MacLarnon & Hewitt, 2004). In summary, humans have a higher risk of choking while eating but are better at producing carefully controlled vocalizations. Therefore, following Parker and Maynard Smith (1990), who observed that optimization for a function often indicates selective pressure related to that function, it seems likely that humans have undergone selective pressure related to speech.

In this opinion piece, I want to address three questions about the evolution of speech and its relation to the evolution of language:

1. How long ago did adaptations for speech evolve?
2. To what extent did speech and language evolve independently?
3. Is anatomy or cognition primary?

The first question is important, because it not only addresses the issue of how long ago hominins started to speak, but also indirectly addresses the wider issue of how much time has been available for the evolution of language. However, the importance of this evidence depends on the second question: Could physical adaptations of the vocal tract to produce more complex and more flexible vocalizations have evolved independently from cognitive

✉ Bart de Boer
bart@ai.vub.ac.be

¹ Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

adaptations (e.g., for ecological reasons: Fitch & Hauser, 2002; Hombert, 2010)? And, even if cognitive adaptations did evolve in parallel, were they really related to language, or could they have been used for something else? This question has two sides: Could complex vocalization have evolved for something other than language (e.g., for singing: Fitch, 2010, chap. 14; Mithen, 2007), and could language have (initially) evolved without complex vocalizations (e.g., through a gestural origin: Corballis, 2002)?

Finally, many conclusions have been based on anatomical evidence for or against the ability to produce complex vocalizations. However, this assumes that (modern) language somehow depends on modern anatomy, and the question is whether this is true.

Paleontological evidence for speech

To answer the question of when adaptations to speech first appeared, it is useful to review the paleontological evidence (Barney, Martelli, Serrurier, & Steele, 2012; Dediu & Levinson, 2013; Fitch, 2009). The oldest and most controversial evidence is about the position of the larynx. Lieberman and Crelin (1971) reconstructed the Neanderthal vocal tract on the basis of observations of fossils and of the infant modern human vocal tract. They then used a computer model to calculate the acoustic properties of articulations that could be made with the reconstructed vocal tract. They concluded that Neanderthals were not able to produce the complete range of sounds that modern humans can make. Although this work was groundbreaking in its method, the results have not been universally accepted, in part because their reconstruction was considered wrong (Arensburg, Schepartz, Tillier, Vandermeersch, & Rak, 1990; Schepartz, 1993), but also because the computer model explored only a small set of articulations (Boë, Heim, Honda, & Maeda, 2002; but see also de Boer & Fitch, 2010). More recent reconstructions and computer models have tended to lead to the conclusion that Neanderthals had articulatory abilities similar to those of modern humans (Boë et al. 2002). The main obstacle preventing consensus is that there is no universally accepted reconstruction of the Neanderthal vocal tract on the basis of the fossils we currently have. Nevertheless, more recent review articles have tended to attribute modern human-like articulatory abilities to Neanderthals (Barney et al. 2012; Dediu & Levinson, 2013).

A recent study (D’Anastasio et al. 2013) circumvented the problems of a direct reconstruction of the vocal tract by looking only at the internal composition of a Neanderthal hyoid bone (Arensburg et al. 1989). Although it was already known that the shape of the Neanderthal hyoid bone falls within the range of modern human possibilities (Arensburg

et al. 1990), D’Anastasio et al. found that in addition, the ways in which muscles attached to it and exerted stresses on it are indistinguishable between modern humans and Neanderthals.

Fossil hyoid bones are important for another reason, because they are associated with the presence of air sacs (Fitch, 2000; Steele, Clegg, & Martelli, 2013, for a recent review). All great apes except modern humans have air sacs, and both acoustic analysis (de Boer, 2009; Riede, Tokuda, Munger, & Thomson, 2008) and experiments (de Boer, 2012) have indicated that their presence would decrease the understandability of vocalizations. Fossil hyoid bones of Neanderthals (Arensburg et al. 1989), *Homo heidelbergensis* (Martínez et al. 2008), and contemporaries (Capasso, Michetti, & D’Anastasio, 2008) indicate that they most likely did not have air sacs, whereas a hyoid fragment (a bulla) of *Australopithecus afarensis* that is shaped like that of a chimpanzee (Alemseged et al. 2006) indicates that Australopithecines most likely did have air sacs. Still, this is only indirect evidence, since air sacs may have disappeared in human evolution for other reasons, such as prevention of hyperventilation during long calls (Hewitt, MacLarnon, & Jones, 2002).

MacLarnon and Hewitt (1999, 2004) have also proposed a different fossil indication of the evolution of speech: the thoracic vertebral canal. They argue that a larger thoracic vertebral canal indicates a larger number of neural fibers, and that this can be used to investigate how well the intercostal muscles are controlled. Better control is needed for the extremely long and accurate outbreaths that are used in speech. These researchers found that the thoracic vertebral canal is relatively larger in modern humans and Neanderthals than in other apes, whereas *Homo ergaster* had a thoracic vertebral canal of the size that would be expected in an ape their size.

Finally, Martínez et al. (2004) proposed that the hearing of *Homo heidelbergensis* was similar to that of modern humans, whereas that of chimpanzees and earlier hominins was different. They interpreted this as an indication that modern human (and Neanderthal) hearing is specialized for speech, because it is more sensitive to frequencies in the range of 2–4 kHz. However, it has to be noted that this is rather higher than what is usually considered essential for speech (300–3300 Hz was considered sufficient for analog telephone speech).

Although none of the individual pieces of evidence are unequivocal, when they are taken together a relatively consistent picture emerges: Both *Homo sapiens* and Neanderthals—and by extension their latest common ancestor, who lived about 400,000 years ago—have the same anatomical adaptations to speech, whereas earlier species were probably different. This indicates that the ability to produce complex vocalizations was already present 400,000 years ago. Moreover, the observed changes are very diverse, but they are all explained by the evolution of complex vocalization.

Does speech imply language?

Even if one accepts that adaptations for complex vocalizations evolved around 400,000 years ago, this does not necessarily mean one accepts that language evolved at that time, as well. Two alternatives are possible: that language (in the sense of a learned symbolic and open-ended communication system) came before speech, or that language came after. The first position requires that language was initially conveyed through some other medium, and the gestural-origins hypothesis proposes just that (Corballis, 2002). The second position requires that complex vocalizations were originally used for a different purpose, and this is what is advocated by the musical-origins hypothesis (Fitch, 2010, chap. 14; Mithen, 2007). Both positions, in my opinion, are unlikely.

It is unlikely that language first evolved in the gestural modality and later shifted to the acoustic modality, because even if the acoustic modality has certain advantages (e.g., one can use it in the dark), evolution tends to stick with solutions that work well but are not necessarily optimal. There is certainly a very good case for the multimodal origins of language (and for language still being multimodal today; see, e.g., Goldin-Meadow & Alibali, 2013), because other apes (and by homology, our latest common ancestor) are more flexible in using gestures than in using vocalizations (Pika, Liebal, Call, & Tomasello, 2005; Pollick & de Waal, 2007), and because their communication is generally multimodal anyway (Kita, Özyürek, Allen, & Ishizuka, 2010). Nevertheless, apes do show some flexibility in vocalization (Crockford, Herbinger, Vigilant, & Boesch, 2004; de Boer & Perlman, 2014; Perlman, Patterson, & Cohn, 2012; Wich et al. 2009). It therefore seems plausible that vocalizations have played a role in language from the beginning, and that this has caused selective pressure on the vocal tract and other systems involved in speech production and processing.

The hypothesis that vocalizations were first used for music (in the form of song) is also problematic, not because music cannot be old or important, but because it seems unlikely that the extra cognitive abilities needed for music would not also directly have given our ancestors an ability for (simple) language. The prelinguistic cognitive abilities of the latest common ancestor with apes, derived from comparisons of ape abilities, consist of the ability to learn a sizable lexicon of form–meaning mappings (800 words in gorillas, Patterson & Cohn, 1990; and chimpanzees, Savage-Rumbaugh, McDonalds, Sevcik, Hopkins, & Rubert, 1986) and the ability to do basic semantics (Hurford, 2007). The innovations needed for music are vocal imitation and vocal control. In addition, if music is to be a social activity (as it appears to be in humans), an ability for music also includes the ability to cooperate, and to do so vocally. It therefore seems more parsimonious to propose that language and music evolved together and are most likely two sides of the same coin than to propose

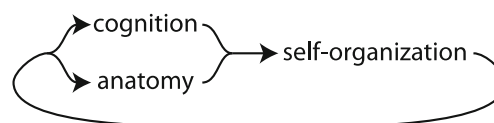
that music evolved much earlier than language. This does not mean that there can be no cognitive specializations for either. The increasing complexity of language and song (under the influence of coevolution; see the next section) could, for instance, lead to linguistic cognitive abilities to deal with complex combinatorial meaning and syntactic structure, as well as musical cognitive abilities to deal with complex rhythm and harmony. Moreover, it should be noted that proponents of musical protolanguage tend to situate the starts of their evolutionary scenarios earlier than 400,000 years ago, usually around the period of *Homo erectus*, so the theory deals with the earliest precursors of language.

These two points indicate that speech and language most likely evolved together, and that we can therefore assume that the beginnings of language are at least 400,000 years old as well.

Coevolution of speech and cognition

A final issue is the relation between cognition and anatomy. Are anatomical innovations the primer for language to arise, or are cognitive adaptations? It is true that anatomical innovations allow for a larger and more fluent set of utterances, and that this puts pressure on the cognitive systems dealing with speech. Also, it has been shown that self-organization under functional constraints can explain, without recourse to specialist cognitive mechanisms, aspects of language such as the structure of vowel and consonant systems (de Boer, 2000; Lindblom, MacNeilage, & Studdert-Kennedy, 1984; Oudeyer, 2005) and the emergence of combinatorial structure (Zuidema & de Boer 2009). On the other hand, it has been pointed out that essential ingredients for speech, such as vocal control and the ability to imitate, are lacking in other apes, and by homology in our latest common ancestor (Ackermann, Hage, & Ziegler, 2014; Fitch, 2010, chaps. 4 and 9).

Anatomy and cognition are expected to coevolve under the influence of self-organization:



When a small anatomical or cognitive innovation occurs, self-organization will cause this to be reflected in the language. This will then change the selective pressure on either the cognitive system or the anatomy, thus creating the potential for further adaptations (de Boer, 2016). For instance, self-organization causes the vowel space to be used maximally, but this causes a selective advantage to speakers with a slightly better articulatory ability. Without the effect of self-organization, such modifications would not have an advantage. Thus, self-organization also helps overcome the problem of the

frequency dependence of the selective advantage of adaptations to language (i.e., what use is an adaptation to language if you are the only one to have it?). Self-organization causes phenomena to emerge in a language that the language users can then adapt to.

The evolution of cognitive and anatomical adaptations is therefore inherently a process of coevolution, and it is misleading to speak about one being more important than the other. Nevertheless, there are arguments for cognitive adaptations being the triggering factor. First, self-organization, in the sense investigated in the articles referred to above, would only be possible if certain cognitive innovations, such as flexibility in vocalization and the ability to do vocal imitation, were already present. In that respect, it is clear that historically, certain cognitive adaptations must have occurred before selective pressure for vocal communication became an issue.

Second, even fully complex modern language is possible without a very large set of speech sounds. For instance, the Rotokas language only uses 11 phonemic contrasts (Firchow & Firchow, 1969). These phonemes are dispersed through the available articulatory space, but this illustrates that a fully modern language does not need many acoustic building blocks; to make 11 phonemic contrasts, only a small subset of the modern articulatory space suffices. An example subset of 11 phonemes of English could be /p/, /b/, /m/, /f/, /v/, /w/, /h/, /i/, / /, /ɛ/, and /æ/. For the consonants, these only make use of bilabial/labiodental and laryngeal articulations (involving the lips or the vocal folds), and for the vowels, only tongue height/jaw opening is relevant. Such articulatory gestures can probably be made by other apes (Lameira, Maddieson, & Zuberbühler, 2014). Although this system would be less robust to noise than most modern phoneme inventories, it would nevertheless be usable, and certainly more useful than no phonological system at all. Hence, even using only the articulatory capabilities of an ape vocal tract, a system with enough contrasts for a full language would be possible.

Conclusion

Convergent evidence for adaptations to complex vocalizations in Neanderthals and *Homo heidelbergensis* indicates that adaptations to producing complex vocalizations were already present 400,000 years ago. In combination with what we know about the prelinguistic abilities of other apes (and thus, of our latest common ancestor), it seems likely that some form of language must have been present as well. Given that even with a monkey-like vocal tract it is probably already possible to produce a range of articulations that is sufficient to be usable for language, it seems that cognitive adaptations must have triggered the emergence of language. However, I certainly do not wish to argue for a catastrophic scenario along the lines of the

minimalist program (Chomsky, 2007). I also do not think the evidence indicates transitions from radically different precursors, such as musical or gestural systems. Rather, the evidence reviewed here indicates a much more gradual process of coevolution between cognition and anatomy; between vocalizations, gesture, and communicative abilities; and between culture and biology, linked through self-organization. It is true that this coevolutionary account does not propose a clear causal factor that may potentially tell us when language emerged precisely, but the complexity of multiple coevolving systems is something that the field of language evolution has been coming to terms with over the last few decades (see also the contributions from the Evolang series of conferences in Cartmill, Roberts, Lyn, & Cornish, 2014). A coevolutionary account is perhaps less spectacular, but it is much more plausible biologically.

References

- Ackermann, H., Hage, S. R., & Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behavioral and Brain Sciences*, 37, 529–546. doi:10.1017/S0140525X13003099
- Alemseged, Z., Spoor, F., Kimbel, W. H., Bobe, R., Geraads, D., Reed, D., & Wynn, J. G. (2006). A juvenile early hominin skeleton from Dikika, Ethiopia. *Nature*, 443, 296–301. doi:10.1038/nature05047
- Arensburg, B., Schepartz, L. A., Tillier, A. M., Vandermeersch, B., & Rak, Y. (1990). A reappraisal of the anatomical basis for speech in middle Palaeolithic hominids. *American Journal of Physical Anthropology*, 83, 137–146.
- Arensburg, B., Tillier, A. M., Vandermeersch, B., Duday, H., Schepartz, L. A., & Rak, Y. (1989). A middle Palaeolithic human hyoid bone. *Nature*, 338, 758–760.
- Badin, P., Boë, L.-J., Sawallis, T. R., & Schwartz, J.-L. (2014). Keep the lips to free the larynx: Comments on de Boer's articulatory model (2010). *Journal of Phonetics*, 46, 161–167. doi:10.1016/j.wocn.2014.07.002
- Barney, A., Martelli, S., Serrurier, A., & Steele, J. (2012). Articulatory capacity of Neanderthals, a very recent and human-like fossil hominin. *Philosophical Transactions of the Royal Society B*, 367, 88–102. doi:10.1098/rstb.2011.0259
- Boë, L.-J., Heim, J.-L., Honda, K., & Maeda, S. (2002). The potential Neandertal vowel space was as large as that of modern humans. *Journal of Phonetics*, 30, 465–484. doi:10.1006/jpho.2002.0170
- Capasso, L., Michetti, E., & D'Anastasio, R. (2008). A *Homo erectus* hyoid bone: Possible implications for the origin of the human capability for speech. *Collegium Antropologicum*, 32, 1007–1011.
- Cartmill, E. A., Roberts, S., Lyn, H., & Cornish, H. (2014). *The evolution of language: Proceedings of the 10th international conference (EVOLANG10)*. Hackensack, NJ: World Scientific.
- Chomsky, N. (2007). Of minds and language. *Biolinguistics*, 1, 9–27.
- Corballis, M. (2002). Did language evolve from manual gestures? In A. Wray (Ed.), *The transition to language: Studies in the evolution of language* (pp. 161–179). Oxford, UK: Oxford University Press.
- Crockford, C., Herbinger, I., Vigilant, L., & Boesch, C. (2004). Wild chimpanzees produce group-specific calls: A case for vocal

- learning? *Ethology*, 110, 221–243. doi:[10.1111/j.1439-0310.2004.00968.x](https://doi.org/10.1111/j.1439-0310.2004.00968.x)
- D'Anastasio, R., Wroe, S., Tuniz, C., Mancini, L., Cesana, D. T., Dreossi, D., ... Capasso, L. (2013). Micro-biomechanics of the Kebara 2 hyoid and its implications for speech in Neanderthals. *PLoS ONE*, 8, e82261:1–7. doi:[10.1371/journal.pone.0082261](https://doi.org/10.1371/journal.pone.0082261)
- de Boer, B. (2000). Self organization in vowel systems. *Journal of Phonetics*, 28, 441–465.
- de Boer, B. (2009). Acoustic analysis of primate air sacs and their effect on vocalization. *Journal of the Acoustical Society of America*, 126, 3329–3343.
- de Boer, B. (2010). Investigating the acoustic effect of the descended larynx with articulatory models. *Journal of Phonetics*, 38, 679–686.
- de Boer, B. (2012). Loss of air sacs improved hominin speech abilities. *Journal of Human Evolution*, 62, 1–6. doi:[10.1016/j.jhevol.2011.07.007](https://doi.org/10.1016/j.jhevol.2011.07.007)
- de Boer, B. (2016). Modeling co-evolution of speech and biology. *Topics in Cognitive Science*, 8, 459–468. doi:[10.1111/tops.12191](https://doi.org/10.1111/tops.12191)
- de Boer, B., & Fitch, W. T. (2010). Computer models of vocal tract evolution: An overview and critique. *Adaptive Behavior*, 18, 36–47.
- de Boer, B., & Perlman, M. (2014). Physical mechanisms may be as important as brain mechanisms in evolution of speech. *Behavioral and Brain Sciences*, 37, 552–553. doi:[10.1017/S0140525X13004007](https://doi.org/10.1017/S0140525X13004007). disc. 577–604.
- Dediu, D., & Levinson, S. C. (2013). On the antiquity of language: The reinterpretation of Neandertal linguistic capacities and its consequences. *Frontier in Psychology*, 4(397), 1–17. doi:[10.3389/fpsyg.2013.00397](https://doi.org/10.3389/fpsyg.2013.00397)
- Firchow, I., & Firchow, J. (1969). An abbreviated phoneme inventory. *Anthropological Linguistics*, 11, 271–276.
- Fitch, W. T. (2000). The evolution of speech: A comparative review. *Trends in Cognitive Sciences*, 4, 258–267. doi:[10.1016/S1364-6613\(00\)01494-7](https://doi.org/10.1016/S1364-6613(00)01494-7)
- Fitch, W. T. (2009). Fossil cues to the evolution of speech. In R. Botha & C. Knight (Eds.), *The cradle of language* (pp. 112–134). Oxford, UK: Oxford University Press.
- Fitch, W. T. (2010). *The evolution of language*. Cambridge, UK: Cambridge University Press.
- Fitch, W. T., & Hauser, M. D. (2002). Unpacking “honesty”: Vertebrate vocal production and the evolution of acoustic signals. In A. M. Simmons, R. R. Fay, & A. N. Popper (Eds.), *Acoustic communication* (pp. 65–137). New York, NY: Springer.
- Goldin-Meadow, S., & Alibali, M. W. (2013). Gesture's role in speaking, learning, and creating language. *Annual Review of Psychology*, 64, 257–283. doi:[10.1146/annurev-psych-113011-143802](https://doi.org/10.1146/annurev-psych-113011-143802)
- Heimlich, H. J. (1975). A life-saving maneuver to prevent food-choking. *Journal of the American Medical Association*, 234, 398–401.
- Hewitt, G. P., MacLamon, A., & Jones, K. E. (2002). The functions of laryngeal air sacs in primates: A new hypothesis. *Folia Primatologica*, 73, 70–94.
- Hombert, J.-M. (2010). Ecological and sexual explanations for larynx lowering. In A. D. M. Smith, M. Schouwstra, B. de Boer, & K. Smith (Eds.), *The evolution of language: Proceedings of the 8th international conference (EVOLANG 8)* (pp. 421–422). Hackensack, NJ: World Scientific.
- Hurford, J. R. (2007). *The origins of meaning: Language in the light of evolution*. Oxford, UK: Oxford University Press.
- Kita, S., Özyürek, A., Allen, S., & Ishizuka, T. (2010). Early links between iconic gestures and sound symbolic words: Evidence for multimodal protolanguage [Abstract]. In A. D. M. Smith, M. Schouwstra, B. de Boer, & K. Smith (Eds.), *The evolution of language: Proceedings of the 8th international conference (EVOLANG 8)* (pp. 429–430). Hackensack, NJ: World Scientific.
- Lameira, A. R., Maddieson, I., & Zuberbühler, K. (2014). Primate feedstock for the evolution of consonants. *Trends in Cognitive Sciences*, 18, 60–62. doi:[10.1016/j.tics.2013.10.013](https://doi.org/10.1016/j.tics.2013.10.013)
- Lieberman, P. H., & Crelin, E. S. (1971). On the speech of Neanderthal man. *Linguistic Inquiry*, 2, 203–222.
- Lindblom, B., MacNeilage, P. F., & Studdert-Kennedy, M. (1984). Self-organizing processes and the explanation of language universals. In M. Butterworth, B. Comrie, & Ö. Dahl (Eds.), *Explanations for language universals* (pp. 181–203). de Gruyter: Berlin, Germany.
- MacLamon, A., & Hewitt, G. P. (1999). The evolution of human speech: The role of enhanced breathing control. *American Journal of Physical Anthropology*, 109, 341–343.
- MacLamon, A., & Hewitt, G. P. (2004). Increased breathing control: Another factor in the evolution of human language. *Evolutionary Anthropology*, 13, 181–197.
- Martínez, I., Arsuaga, J.-L., Quam, R., Carretero, J.-M., Gracia, A., & Rodríguez, L. (2008). Human hyoid bones from the middle Pleistocene site of the Sima de los Huesos (Sierra de Atapuerca, Spain). *Journal of Human Evolution*, 54, 118–124. doi:[10.1016/j.jhevol.2007.07.006](https://doi.org/10.1016/j.jhevol.2007.07.006)
- Martínez, I., Rosa, M., Arsuaga, J.-L., Jarabo, P., Quam, R., Lorenzo, C., ... Carbonell, E. (2004). Auditory capacities in Middle Pleistocene humans from the Sierra de Atapuerca in Spain. *Proceedings of the National Academy of Sciences*, 101, 9976–9981.
- Mithen, S. (2007). *The singing Neanderthals: The origins of music, language, mind and body*. Cambridge, MA: Harvard University Press.
- Oudeyer, P.-Y. (2005). The self-organization of speech sounds. *Journal of Theoretical Biology*, 233, 435–449.
- Parker, G. A., & Maynard Smith, J. (1990). Optimality theory in evolutionary biology. *Nature*, 348, 27–33. doi:[10.1038/348027a0](https://doi.org/10.1038/348027a0)
- Patterson, F. G., & Cohn, R. H. (1990). Language acquisition by a lowland gorilla: Koko's first ten years of vocabulary development. *Word*, 41, 97–143.
- Perlman, M., Patterson, F. G., & Cohn, R. H. (2012). The human-fostered gorilla Koko shows breath control in play with wind instruments. *Biological Sciences*, 6, 433–444.
- Pika, S., Liebal, K., Call, J., & Tomasello, M. (2005). The gestural communication of apes. *Gesture*, 5, 41–56.
- Pollick, A. S., & de Waal, F. B. M. (2007). Ape gestures and language evolution. *Proceedings of the National Academy of Sciences*, 104, 8184–8189. doi:[10.1073/pnas.0702624104](https://doi.org/10.1073/pnas.0702624104)
- Riede, T., Tokuda, I. T., Munger, J. B., & Thomson, S. L. (2008). Mammalian laryngeal air sacs add variability to the vocal tract impedance: Physical and computational modeling. *Journal of the Acoustical Society of America*, 124, 634–647.
- Savage-Rumbaugh, S., McDonalds, K., Sevcik, R. A., Hopkins, W. D., & Rubert, E. (1986). Spontaneous symbol acquisition and communicative use by pygmy chimpanzees (*Pan paniscus*). *Journal of Experimental Psychology*, 115, 211–235.
- Schepartz, L. A. (1993). Language and modern human origins. *Yearbook of Physical Anthropology*, 36, 91–126.
- Steele, J., Clegg, M., & Martelli, S. (2013). Comparative morphology of the hominin and African ape hyoid bone, a possible marker of the evolution of speech. *Human Biology*, 85, 639–672. doi:[10.3378/027.085.0501](https://doi.org/10.3378/027.085.0501)
- Wich, S. A., Swartz, K. B., Hardus, M. E., Lameira, A. R., Stromberg, E., & Shumaker, R. W. (2009). A case of spontaneous acquisition of a human sound by an orangutan. *Primates*, 50, 56–64. doi:[10.1007/s10329-008-0117-y](https://doi.org/10.1007/s10329-008-0117-y)
- Zuidema, W., & de Boer, B. (2009). The evolution of combinatorial phonology. *Journal of Phonetics*, 37, 125–144. doi:[10.1016/j.wocn.2008.10.003](https://doi.org/10.1016/j.wocn.2008.10.003)