The Evolution of Speech

Human speech – that is, the acoustic signal used for language – differs in a number of respects from the sounds produced by our nearest relatives, the great apes. Like apes and other primates, humans have a system to convey very basic information about emotional state using signals such as crying, laughter and cries of pain or joy. However, parallel to this system, humans use speech to convey complex abstract information. There are two fundamental differences between the basic primate call system and speech. First of all, speech is under voluntary control, whereas the basic call system is almost completely involuntary. Secondly, the basic call system is holistic, whereas speech is combinatorial. A combinatorial system contains a small number of basic elements that can be combined into many more different sequences. In the case of speech, the basic elements are phonemes or syllables, and these are combined into words and phrases. Holistic utterances, on the other hand do not consist of smaller elements. There is no meaningful way in which they can be recombined, although they might have some parameters that can be varied, such as pitch and loudness.

Humans have several adaptations to using an acoustic communication system that is under voluntary control and that uses recombination of a potentially large number of basic units. Humans have extraordinarily good control over breathing, especially on the outbreath. Also, humans have a vocal tract that is adapted to producing a large number of distinct sounds. In contrast with other mammals, whose vocal tracts can be considered as a single tube, the human vocal tract consists of two linked tubes. These are the pharyngeal cavity and the oral cavity. Two linked tubes allow a much larger repertoire of possible sounds than a single tube. The body of the tongue makes the division between the two tubes. As the human tongue is so important in controlling articulation, and as it needs to move rapidly for producing speech, it is relatively small compared to the tongues of other primates, and extremely well innervated. Human hearing is also adapted to speech. Humans are very sensitive to sounds between 1kHz and 4kHz. This range is very important for speech. It is approximately the range of the important resonance frequencies of the human vocal tract. These resonance frequencies (formants) determine the acoustic character of a speech sound.

Humans also have many neurological adaptations to speech, although these are harder to measure. It appears that humans are extremely good at performing rapid controlled and coordinated movements with both their hands and their articulatory organs. In this respect humans are capable of more accurate motor control and planning than other primates. Interestingly, rapid hand movements are also used in human sign language, used mostly by deaf people who are unable to communicate acoustically.

All these properties must be explained by an evolutionary account of speech. The next sections will provide background information on the different sources of evidence for the evolution of speech, as well as an example evolutionary scenario.

Fossil evidence

As speech is the most physical aspect of language, it is the most likely to leave direct traces in the fossil record. Unfortunately, the shape of the human vocal tract is determined mostly by soft tissue. It is therefore hard if not impossible to reconstruct its evolution from the fossils that we have. However, there are other adaptations to speech, such as better control of breathing and of tongue movements. As larger nerves are needed for improved control, some researchers have investigated the sizes of the holes in the skull and vertebrae through which these nerves pass.

Kay, Cartmil and Balow (1998) have investigated the size of the hypoglossal canal. This is the canal through which the main nerve that innervates the tongue passes. They determined that it is larger (both absolutely and relatively) in modern humans than in modern chimpanzees and gorillas. They then compared the sizes of the hypoglossal canal in different fossils and found that those of modern and early *Homo sapiens* as well as those of Neanderthals fall within the modern human range, while those of *Australopithecus* and *Homo habilis* fall within the chimpanzee range. From this they conclude that "human vocal abilities" were present about 400 000 years ago.

MacLarnon and Hewitt (1999) have compared the thoracic vertebral canal of modern humans, apes and fossils of human ancestors. The thoracic vertebral canal is the canal through which the nerve for controlling the diaphragm, and therefore breathing, passes. They have found that it is significantly larger with respect to body weight in modern humans than in apes, and that Neanderthals and early *Homo sapiens* fall in the modern human range, while *Homo ergaster* and the Australopithecines do not. They provide a somewhat larger range for the origin of increased breathing control: between 1.6 million and 100 000 years ago.

Another line of evidence is related to perception of speech sounds. Martínez and colleagues (2004) have calculated the frequency sensitivity of the hearing system of *Homo heidelbergensis* (related to Neanderthals). Their fossils are estimated to be about 350 000 years old. They have found that the frequency sensitivity is comparable to that of modern humans in the range of frequencies that is important for speech. This is in contrast with the hearing system of chimpanzees, which has quite different frequency sensitivity than that of humans.

The above-mentioned evidence seems to suggest that adaptations for speech must have first appeared in the common ancestor of *Homo sapiens* and *Homo neanderthalensis* and that complex speech must have had at least 400 000 years to evolve. Early work by Lieberman and Crelin (1971) however, suggested that the vocal tract of the Neanderthal cannot have consisted of a two-tube system like the modern human vocal tract, giving them a smaller range of possible sounds. This used to be interpreted as an argument against Neanderthal speech, but more recently Lieberman has accepted the possibility of speech in Neanderthals.

Parallels in other animals

In order to get an idea of what the starting point of the evolution of speech was, it is necessary to investigate the vocal abilities of our nearest evolutionary relatives, the apes and other primates. There are also other animals that use complex systems of vocal communication resembling human speech in the sense that they are combinatorial and learned. Examples of these are the songs of certain songbirds and the songs of humpback whales. As these animals are quite distant evolutionary, and as there are many intermediate species that do not have such vocal abilities, they must be considered interesting examples of convergent evolution. However, they provide evidence that acoustic, learned and combinatorial communication systems are useful for other species. By comparing relates species, one can get an idea of how these systems have evolved. Fitch (2000) provides an overview of other animals' vocalizations in the context of the evolution of speech.

Apes and other primates do use vocal communication extensively, but as has been mentioned above, primate vocalizations are more akin to human emotional utterances than to speech. Examples of research cited by Hauser (1997) indicate that primate vocalizations are mostly innate, but that details of their production might be learned. He also indicates that there has been "no convincing demonstration" of imitation of vocal utterances except in captive chimpanzees and orangutans. There have been several attempts to teach orangutans or chimpanzees to speak, all of which have failed compared to attempts to teach apes sign language. This is perhaps not very surprising, given that apes have a very different vocal tract, and also have much less accurate motor control. However, the attempts do show that they have at least some voluntary control over vocalization. As for perception, apes are able to understand spoken utterances to some degree, and therefore must be able to perceive at least some of the acoustic distinctions made in human language.

Surprisingly little is known about the structure of primate calls and how it compares to human phonology. Of the utterances of the apes, the long calls of the different gibbon species are perhaps best studied, and these appear to have compositional structure, in that smaller units can be recombined into novel utterances (Mitani & Marler, 1989). Among the apes, gibbons are most distantly related to humans, and all more closely related species have less complex calls. Therefore, gibbon calls cannot be considered direct precursors of human speech. Nevertheless they illustrate two points: that combinatorial call systems have emerged in other contexts, and that higher primate brains appear to be able to handle the complexity of producing and processing combinatorial calls.

There is considerable anatomical overlap between the organization of vocalizations in the primate brain and speech in the human brain. According to Ploog (2002), there are two pathways in both humans and other primates: the "cingulate vocalisation pathway" mostly involved with emotional content of speech, and the "neocortical vocal pathway" involved with fine motor control. The latter appears to be more developed in species that are more closely related to humans.

Mechanisms

The above mentioned observations make clear that (proto-) humans had adaptations for modern speech about 400 000 years ago, and that the last common ancestor of chimpanzees and humans probably already had important pre-adaptations for speech, but was incapable of articulated speech itself. It had more limited articulatory possibilities, as well as insufficiently accurate control of articulation and breathing. It was probably also unable to imitate and learn utterances on any large scale, but possibly had the ability to this on a very limited scale. This could be considered the starting point of the evolution of speech.

It is interesting to consider modern human speech to get an idea of what abilities have evolved. Modern human speech can be amazingly complex. Maddieson (1984) and (Ladefoged & Maddieson, 1996) give an overview of the repertoires of speech sounds (and syllables) that are used in human languages. These can be extremely complex. However, for investigating the evolution of speech and language it is useful to know that there are also languages with very simple sound systems. The simplest sound system is that found in Piraha, which uses only 11 phonemes, while Hawaiian is one of the languages with the smallest number of different syllables: 162 according to (Maddieson, 1984). It is also sometimes said that modern language is not possible without the vowels /i/, /a/ and /u/. However, there are languages with so-called vertical vowel systems, which only contain central vowels. Apparently, it is not necessary to have hugely complex speech for modern language.

There are several mechanisms that need to be taken into account when constructing a theory about the evolution of speech. First of all, speech is a learned system for communication in a population, and therefore cultural evolution is important. A genetic change that would allow an individual to produce or perceive signals that cannot be used by other individuals in the population would be useless. Therefore, cultural evolution must partly drive genetic evolution. Innovations in speech spread culturally, after which genetic adaptation to these innovations becomes useful. This is an example of the Baldwin effect.

The fact that speech is a physical signal that is used in a population also causes different kinds of functional pressures to play a role in its evolution. Not all properties of speech need to be explained as the result of adaptations of the brain. Several researchers (e.g. Carré & Mrayati, 1995; Lindblom, MacNeilage, & Studdert-Kennedy, 1984; Stevens, 1972) have proposed different functional pressures that can explain properties of speech. These can be pressures for such things as acoustic distinctiveness, articulatory ease or articulatory stability. Through self-organization in a population, such functional pressures can determine the kinds of sounds that will end up being used in human speech.

Another mechanism is exaptation. Evolution tends to adapt existing systems for new functions. From anatomical parallels between primates and humans, it is clear that many parts of the primate brain, associated with vocalization and motor control, as well as acoustic perception have been exapted for speech. But behaviors must have been exapted as well. MacNeilage and Davis (2000) have proposed the theory that oscillatory movement of the jaw, such as used in chewing forms the basis of the syllable structure of speech.

The theory that speech has evolved from sign language is also an example of exaptation. Although it is remarkable that language can be expressed with both gestures and sounds, and although it is true that parts of the brain are shared between complex gestures and speech, the theory does not explain how and for what reason sign language has evolved into spoken language.

A possible scenario

The preceding sections have presented facts that are pertinent to the evolution of speech, either from fossil evidence, from what we know about our evolutionary relatives or from what we know about the evolutionary process itself. Any attempt to build a scenario of the evolution of language must however be speculative. Such a scenario can nevertheless be useful to illustrate how the different pieces of evidence can fit together. It must however be seen as nothing more than a possible scenario. There is as yet insufficient consensus amongst researchers to present anything like a definitive scenario.

Evolution of speech must have started with the ability to learn and imitate utterances, so that a culturally transmitted repertoire of sounds could emerge. This repertoire could at first also have consisted of a combination of gestures and sounds. The ability to learn and produce complex articulations could have been exapted from tool use. As vocal communication appears to be the preferred form in primates, the burden of communication in a hybrid system would soon have shifted to vocal communication. When learning of new utterances became possible, it would have become advantageous to make clearer and more distinctive utterances, thus putting evolutionary pressure on breathing control, motor control of articulation and the shape of the vocal tract. At this point one expects to see anatomical adaptations appear in the fossil record. As the repertoire expanded, it became more difficult to learn the different possible utterances. Having utterances with recurring (combinatorial) structure would become advantageous. Any adaptations that made it easier to deal with this structure would be an advantage, although at first it would still be possible to learn the utterances without these adaptations. In a later stage both cognitive abilities and linguistic complexity would increase, until it became impossible to learn to speak without being able to handle combinatorial structure. At this stage, speech had all essential characteristics of modern human speech, even though the possible repertoire of speech sounds was probably much smaller.

Through functional pressure and self-organization in the population, the sound systems that were used would get certain universal characteristics. Certain sounds and combinations of sounds would occur more often than others. It is probable that neural adaptations to these universals have occurred. This is another example of cultural innovation driving genetic evolution. As a result early combinatorial speech evolved into complex modern speech, with all its phonological and phonotactic rules.

Given the fossil evidence, it is clear that early *Homo sapiens* must already have been able of complex articulation, whereas *Homo erectus* probably was not. Whether these articulations were combinatorial remains an open question, but given that gibbons (with much smaller brains) are capable of some degree of recombination in their calls, it is not unlikely that early *Homo sapiens* was able to use combinatorial speech as well. These repertoires of sounds might not have been as complex as those of most modern languages. However, given that there exist modern languages with extremely simple sound systems, there is no reason *from the evolution of speech* that early Homo sapiens was unable of complex communication.

Further reading

Carré, R., & Mrayati, M. (1995). Vowel transitions, vowel systems, and the Distinctive Region Model. In C. e. a. Sorin (Ed.), *Levels in Speech Communication: Relations and Interactions* (pp. 73–89). Amsterdam: Elsevier.

Fitch, W. T. (2000). The evolution of speech: a comparative review. *Trends in cognitve science, 4*(7), 258–267.

Hauser, M. D. (1997). *The evolution of communication*. Cambridge, MA: MIT Press.

Kay, R. F., Cartmill, M., & Balow, M. (1998). The hypoglossal canal and the origin of human vocal behavior. *Proceedings of the National Academy of Sciences, 95*, 5417–5419.

Ladefoged, P., & Maddieson, I. (1996). *The Sounds of the World's Languages*. Oxford: Blackwell.

Lieberman, P., & Crelin, E. S. (1971). On the speech of Neanderthal man. *Linguistic Inquiry, 2*, 203–222.

Lindblom, B., MacNeilage, P., & 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). Berlin: Walter de Gruyter & Co.

MacLarnon, A., & Hewitt, G. P. (1999). The Evolution of Human Speech: the role of enhanced breathing control. *American Journal of Physical Anthropology, 109*(3), 341–343.

MacNeilage, P. F., & Davis, B. L. (2000). On the Origin of Internal Structure of Word Forms. *Science, 288*, 527–531.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

Martínez, I., Rosa, M., Arsuaga, J.-L., Jarabo, P., Quam, R., Lorenzo, C., et al. (2004). Auditory capacities in Middle Pleistocene humans from the Sierra de Atapuerca in Spain. *Proceedings of the National Academy of Sciences, 101*(27), 9976–9981.

Mitani, J. C., & Marler, P. (1989). A phonological analysis of male gibbon singing behavior. *Behaviour, 109*, 20–45.

Ploog, D. (2002). The neural basis of vocalization. In T. J. Crow (Ed.), *The speciation of modern Homo sapiens* (pp. 121–135). Oxford: Oxford University Press.

Stevens, K. N. (1972). The Quantal Nature of Speech: Evidence from articulatory-acoustic data. In E. E. J. David & P. B. Denes (Eds.), *Human communication: a unified view* (pp. 51–66). New York: McGraw-Hill.