

Is the Neural Basis of Vocalisation Different in Non-Human Primates and *Homo sapiens*?

DETLEV PLOOG

Summary. From an evolutionary perspective the voice was a prerequisite for the emergence of speech. Speech, the most advanced mode of vocal communication, became possible only after gradual transformations of the sound-producing system and its central nervous control, in co-evolution with the transformations of the auditory system, had taken place. There are two systems in the brain that produce and control vocal behaviour. The first is very old phylogenetically. In non-human primates and humans it comprises limbic structures, all of which funnel into the peri-aqueductal grey of the midbrain. If this matrix is destroyed all land-living vertebrates become mute. The second system, the neocortical voice pathway as part of the pyramidal tract, emerged in non-human primates and developed in substance from monkeys to humans; it is indispensable for the voluntary control of the voice. The destruction of this system has no influence on the monkey's spontaneous vocal behaviour; in humans, however, it has disastrous consequences and makes speech impossible. The hypothesis is advanced that the last step in the evolution of the phonatory system in the brain was the outgrowing and augmenting of the fine fibre portion of the pyramidal tract synapsing directly with the motor nuclei for the vocal cords and the tongue, so that the direct and voluntary control of vocal behaviour became possible. The question raised in my title, 'Is the neural basis of vocalisation different in primates and *Homo sapiens*', must, of course, be answered with 'yes'. The neural basis is in fact quite different. Explaining this difference and its consequences for the evolution of language and speech is the purpose of this chapter.

INTRODUCTION

VOCAL BEHAVIOUR is a prerequisite for speech behaviour. It has a very long evolutionary history in vertebrates, whilst speech behaviour has a short one. If we follow the evolutionary history of vocal behaviour from toads and frogs to reptiles and mammals, not to mention birds, we can observe profound changes within the sound-producing apparatus, from the tripartite to the quadripartite larynx, including the well-known descent of the human larynx during the first months of ontogenesis. That this transformation is necessary for the development of intelligible speech is demonstrated by children with Down's syndrome, where this descent is incomplete. The genetically determined malfunction impairs speech considerably.

Changes in the internal laryngeal voice-producing muscles and other evolutionary transformations of the larynx should also be considered here. Only in the human species does the vocal muscle (*m. thyreoarytaenoideus lateralis*) send fine fibres into the medial part of the vocal cords, which allows extremely fine tuning of the cords, a prerequisite for the human faculty of singing.

Each little step in the evolution of the voice-producing system led to a higher complexity of vocal behaviour, resulting in different species-specific vocal repertoires, which are used almost exclusively for sexual selection and social communication. Even genetically determined slight differences in the vocal expressions of a given species, so-called 'dialects', have selective consequences (Ploog *et al.*, 1975). I am convinced that vocal gestures are at the roots of the evolution of language, and not body gestures, as Michael Corballis and others believe. Audio-vocal behaviour as opposed to gestural behaviour is advantageous in the dark, and also while hunting, harvesting, cooking, making tools and performing many other daily routines during which communication takes place while the hands are busy. The chief argument, however, is the co-evolution of the vocal and auditory systems, which is rarely mentioned in this debate. Therefore, before I come to the neural basis of vocalisation in non-human primates and humans I will briefly comment on the auditory part of the communication system (Ploog, 1981, 1988, 1990).

AUDIO-VOCAL SIGNALLING

Very early in the evolution of the vertebrates audio-vocal signals function in sexual selection. Female frogs, for example, distinguish between the mating croaks of mature and immature male frogs, and they prefer the voices of their own population to those of neighbouring populations. Moreover, the evolution of the ear from reptiles to mammals, and especially primates, is striking. While reptiles have only one bone for sound transmission in the middle ear,

mammals have three, which required substantial bony transformations of the lower jaw. Why is it that the sound-producing system and the sound-decoding system have co-evolved over 200 or more million years? The answer is that audio-vocal behaviour was the most successful type of behaviour in the increasingly complex sexual selection processes and social communication. And finally, why is it that the new-born human baby is capable of distinguishing certain phonemes from others that are all universal in the languages of *Homo sapiens*? In my opinion, the co-evolution of auditory and vocal behaviour, including the prominent evolution of the brain structures involved, forms the basis of the evolution of language.

TWO SYSTEMS FOR VOCAL BEHAVIOUR

The remainder of my chapter deals with the brain structures involved in vocalisation. My colleagues and I investigated the central nervous organisation of phonation in the squirrel monkey, a small South American primate that is endowed with a rich vocal repertoire. We found that the great variety of calls predominantly regulates the complex social behaviour of these monkeys. Only a few calls refer to external events, especially and differentially to aerial and terrestrial predators. All call types are only slightly modifiable fixed-action patterns, which means that they are innate (Winter *et al.* 1966, 1973; Herzog & Hopf, 1984). There are at least two subspecies with slightly different vocal repertoires that are genetically transmitted (Ploog, 1986, 1995). Our findings in squirrel monkeys are also basically relevant for macaques and other primate species.

Most, if not all, natural vocal expressions can be elicited reliably and repeatedly by electrical stimulation of specific brain sites in the awake animal (Jürgens & Ploog, 1970). In addition, chemical stimulation and pharmacological blockade of specific brain sites have been used to explore further functional properties of the vocal system in the brain (Jürgens & Lu, 1993).

Figure 1 gives an overview of those primate brain structures from which species-specific vocalisations can be elicited electrically (indicated in black). Only limbic, thalamic, hypothalamic and brainstem structures, but no neocortical structures, are part of the extended system. The frontal stippled area is involved in a rather limited control of the voice in simple vocal conditioning paradigms. The stippled area in the peri-aqueductal grey (PAG) of the mid-brain delineates the matrix of the vocal system. Its destruction results in mutism in land-living vertebrates, including humans (Jürgens & Ploog, 1970, 1976).

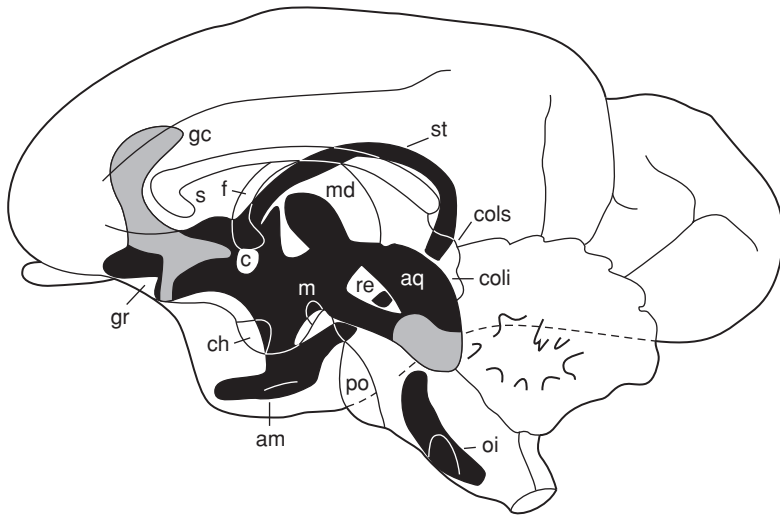


Figure 1. Vocalisation-producing brain areas in the squirrel monkey (*Saimiri sciureus*) (Jürgens & Ploog, 1976). Key: am, amygdala; aq, substantia grisea centralis; c, commissura anterior; ch, chiasma opticum; coli, colliculus inferior; cols, colliculus superior; f, fornix; gc, gyrus cinguli; gr, gyrus rectus; m, corpus mammillare; md, nucleus medialis dorsalis thalami; oi, nucleus olivaris inferior; po, griseum pontis; re, formatio reticularis tegmenti; s, septum; st, stria terminalis.

The cingulate vocalisation pathway

Combined lesion and tracer studies show that all the limbic structures from which vocalisations are elicitable converge in the PAG (Figure 2). There is one major pathway, the 'cingulate vocalisation pathway' (Jürgens & Pratt, 1979), that runs from the anterior cingulate gyrus monosynaptically into the PAG. Lesions along this pathway abolish calls elicited from the anterior cingulate. The course of this tract joins the pyramidal tract in the internal capsule and follows it down to the caudal diencephalon. On leaving the pyramidal tract at the cerebral peduncle, the fibres ascend dorsally to the PAG and follow its course to its end, where they sweep laterally through the parabrachial area and descend through the lateral pons and lateral medulla oblongata to the nucleus ambiguus, which is the nucleus for the internal laryngeal motor neurones (Müller-Preuss & Jürgens, 1976). These anatomical details are important for conclusions to be drawn later.

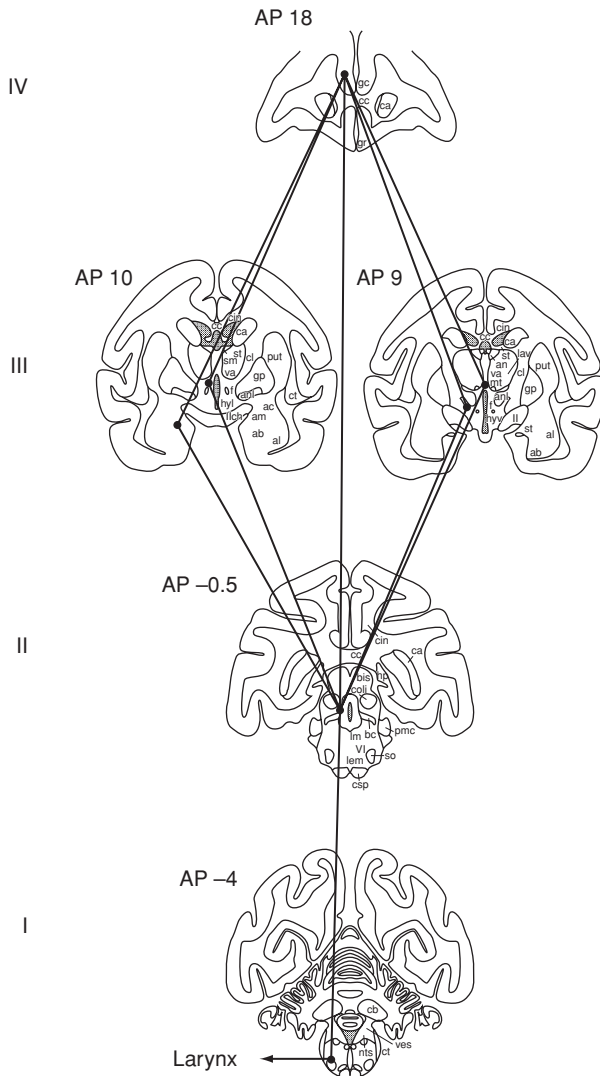


Figure 2. Scheme of hierarchical control of vocalisation. All brain areas indicated by a dot yield vocalisation when electrically stimulated. All lines interconnecting the dots represent anatomically verified direct projections (leading in rostrocaudal direction). The dots indicate in (IV) the anterior cingulate gyrus, in (III) the basal amygdaloid nucleus, dorsomedial and lateral hypothalamus and midline thalamus, in (II) the peri-aqueductal grey and laterally bordering tegmentum, and in (I) the nucleus ambiguus and surrounding reticular formation (the nucleus ambiguus itself only yields isolated movements of the vocal folds; phonation can be obtained, however, from its immediate vicinity). For further information, see the text (Jürgens & Ploog, 1981).

The cortical larynx area and its projection

Electrical stimulation of the most rostrinferior part of the neocortical face representation in the monkey (corresponding to area 6ba in the Brodmann–Vogt nomenclature) yields movements of the vocal folds. Bilateral removal of this area leaves the spontaneous vocalisations of both rhesus and squirrel monkeys unimpaired (Sutton *et al.*, 1974; Kirzinger & Jürgens, 1982). In humans, however, damage to the cortical face area in the dominant hemisphere results in a complete bilateral paralysis of the vocal cords. Such aphonia may last for months (Jürgens *et al.*, 1982). This difference between monkeys and humans can be explained by the fact that in humans, and apparently only in humans, there is a direct pathway from the laryngeal representation in the primary motor cortex to the laryngeal motor neurones of the nucleus ambiguus (Kuypers, 1958).

To clarify further the function of this part of the corticobulbar tract, Jürgens & Zwirner (1996) performed a crucial experiment, in which the PAG was pharmacologically blocked. In this condition it was not possible to elicit vocalisations from the anterior cingulate cortex. Different vocalisation sites in the forebrain could be blocked by a single PAG injection of kynurenic acid or procaine. However, that blockade had no effect on vocal fold movements elicited electrically from the cortical face area.

This finding strongly suggests the existence of two separate pathways controlling the vocal cords, namely the phylogenetically old limbic cingular vocalisation pathway and the phylogenetically young corticobulbar pathway, which is part of the pyramidal tract, a structure found only in mammals. Its entire spinal component is best developed in primates and reaches its greatest development in humans. The most important evolutionary step in its development is a group of corticospinal fibres that originate in the precentral cortex (area 4) and project monosynaptically to motor neural pools in the most ventral part of the cervical and lumbar enlargements of the spinal cord.

The neocortical vocal pathway

This corticomotoneuronal pathway emerges in mammals and develops in substance from monkeys to apes to humans, executing highly fractionated movements of the hands and relatively independent finger movements. This pathway is indispensable for the voluntary control of movements (Hepp-Reymond, 1988; Phillips, 1979). It does not seem too far-fetched to assume that this growing tendency of increasing dexterity and control in evolution is also effective in regard to the fine tuning of the vocal folds by the laryngeal motoneurones, for example in human songs, and the hypoglossal neurones for lingual articulation, which is crucial for speech but seems to be almost completely absent in the

monkey's vocal communication. Again Jürgens and his group did some pioneering work in this area (Chen & Jürgens, 1995). They compared the projections of the tongue area of the primary motor cortex in the tree shrew, the tamarin and the rhesus monkey. They found that the tree shrew lacks a direct connection between motor cortex and hypoglossal nucleus. In the tamarin very few fibres could be detected there, and in the rhesus monkey there were marked terminal fibres in the hypoglossal nucleus. The authors concluded that 'there is a phylogenetic trend from lower to higher primates strengthening the cortico-hypoglossal connections'.

THE HUMAN CASE

In humans, the neocorticobulbar (pyramidal) system and the limbic vocal system co-operate inseparably, but may be separately involved in certain clinical cases.

The cingulate gyrus

A 41-year-old male patient was seen after a cerebral infarction, which was found to have affected the anterior cingulate cortex bilaterally, the left supplementary motor area, and the medial orbital cortex bilaterally (Figure 3). During the first 6 weeks post-infarction the patient remained in a state of akinetic mutism. Occasionally he responded to a painful stimulus with a moan. After 10 weeks, he could repeat long sentences in the whispering mode, articulating clearly without mistakes. Spontaneous utterances, however, remained reduced to a few monosyllabic words. After a year his speech was still restricted and monotonous. In an intonation test performed after 5 years, he was unable to attach the appropriate emotion to certain short exclamations, such as 'shut up' or 'terrific!'. The sonograms revealed that his ability to speak with emotion was greatly reduced, and he was unable to correct this deficiency voluntarily. Responsible for this lack of prosody and the inability to control intonation was the bilateral lesion of the anterior cingulate cortex (Jürgens & Cramon, 1982; Cramon & Jürgens, 1983).

The cortical face area

A second clinical example demonstrates that in humans that part of the cortex in which larynx, pharynx, tongue and mouth are represented is necessary for controlling the voice, whereas in the monkey these structures are not needed for phonation. A 52-year-old right-handed male patient had had an embolic cerebral infarction of the left middle cerebral artery (Figure 4a). He had

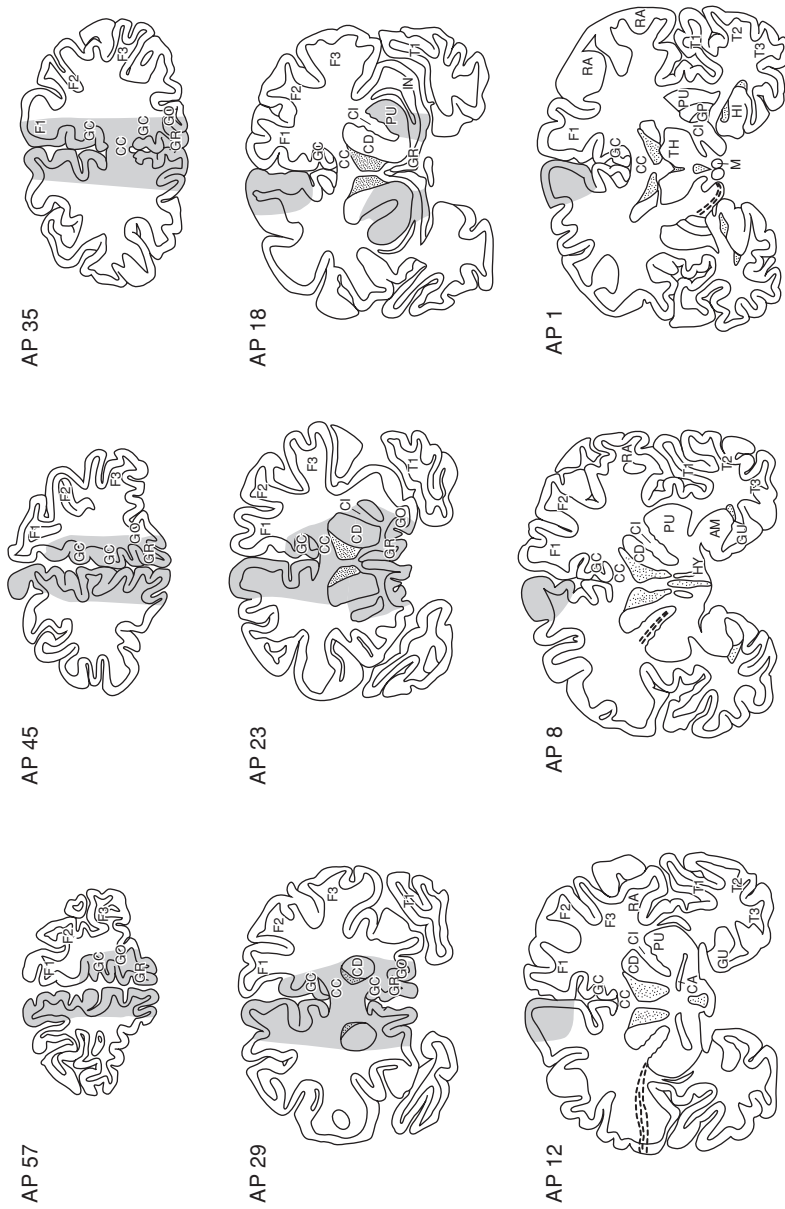


Figure 3. Schematic view of the human brain at nine levels from front (AP 57) to rear (AP 1). Dark zones: cerebral infarction sites, AP 12, AP 8, AP 1. Dashed lines: motor pathway from cortex to larynx and articulatory organs (Jürgens & Cramon, 1982).

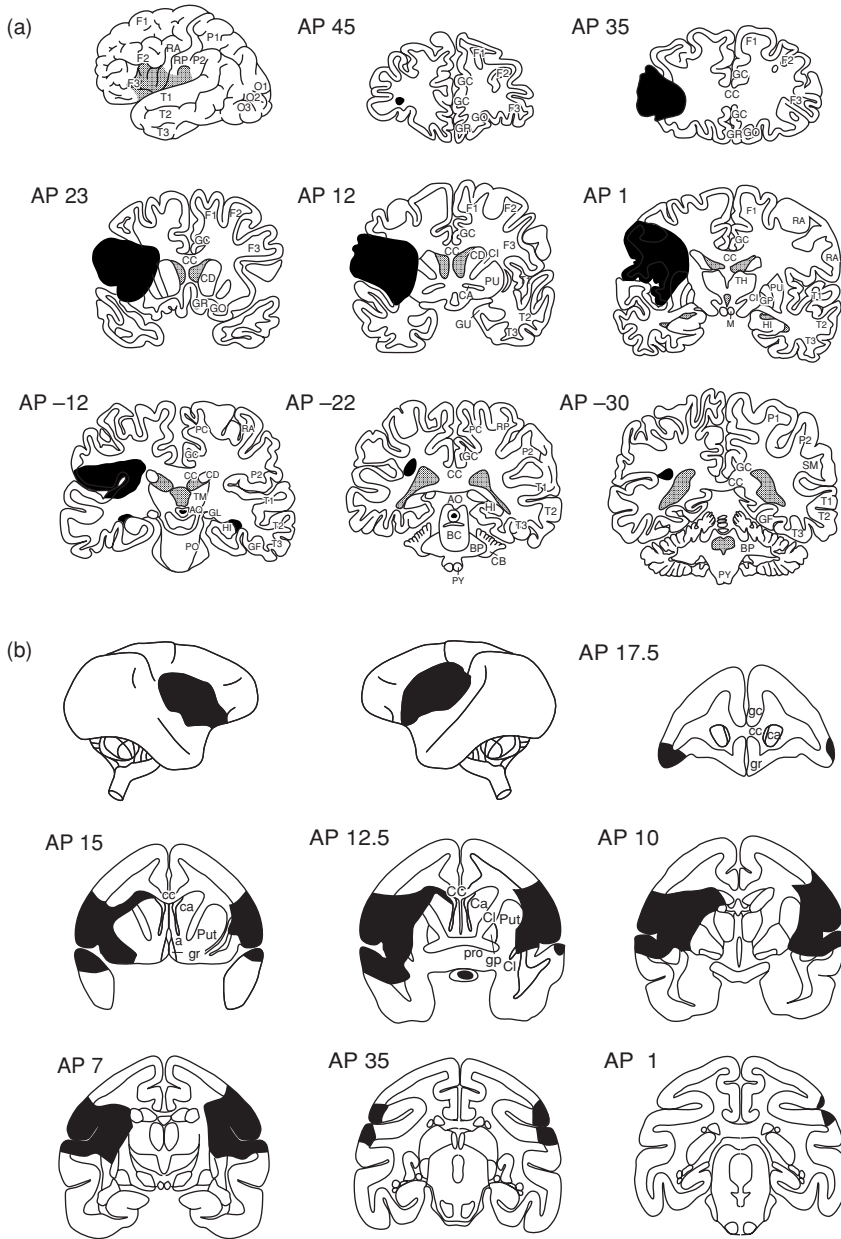


Figure 4. Frontal sections through (a) human brain and (b) monkey brain. Dark zones: (a) cerebral infarction area, (b) cerebral lesion sites (Jürgens *et al.*, 1982).

hemiplegia on the right side, a right-sided lower facial paresis, and tongue deviation to the right. After regaining consciousness, he could understand what was said to him but could not utter a single sound except that he coughed when the base of his tongue was touched. A laryngoscopic examination revealed that his vocal folds remained motionless during both respiration and attempts at phonation. This state of complete mutism and inability to phonate lasted for 11 weeks. Then, in the subsequent 2 weeks, phonation was completely restored. In contrast to the patient in the first example, this patient had considerable difficulty articulating and making oral movements.

The patient's brain lesion was experimentally modelled in the monkey (Figure 4b). In non-human primates vocalisations are elicitable from subcortical areas either in the left or in the right hemisphere. Therefore, the lesion in the monkey was carried out bilaterally. Homologous to the patient's lesion, it invaded the equivalent of Broca's area, the inferior pre- and post-central cortex, the rolandic operculum, and other structures. Although the monkey's tongue, lips and masticatory muscles were completely paralysed after the operation, its phonation remained intact. The spectrographically recorded vocalisations included all call types of the species. Consequently, the vocal folds were functioning, i.e. the central nervous patterning of calls was not impaired (Jürgens *et al.*, 1982).

The results are summarised in Figure 5. On the left side we see the limbic pathway, running from the cingulate gyrus monosynaptically to the PAG (AP -0.5), and from there to the laryngeal motor neurone (AP -4); on the right side, the corticopyramidal pathway, running from the cortical face area to the laryngeal motor neurones (AP -4). Triangles indicate sites the electrical stimulation of which produces vocalisation (AP 19) and isolated vocal fold movements, respectively. The dot at AP -0.5 indicates an injection site capable of blocking limbically, but not neocortically, induced vocal fold activity (Jürgens & Zwirner, 1996).

The two systems in tandem

It remains to be explained how these two systems, the neocortical executive system and the limbic vocal system, function inseparably in everyday life. From the neuroanatomical point of view, in the squirrel monkey there are numerous connections between the two systems, of which only a few shall be mentioned here. There are two areas that receive direct projections from the cortical larynx area, namely the anterior cingulate area and the parabrachial nuclei at the PAG, two nodal areas in the limbic vocal system (Jürgens, 1976). Conversely, projections from the limbic cingulate area into neocortical areas, for example the dorsal medial frontal cortex and Broca's area 44, are manifold (Müller-Preuss & Jürgens, 1976). The point is that the limbic part of the vocal

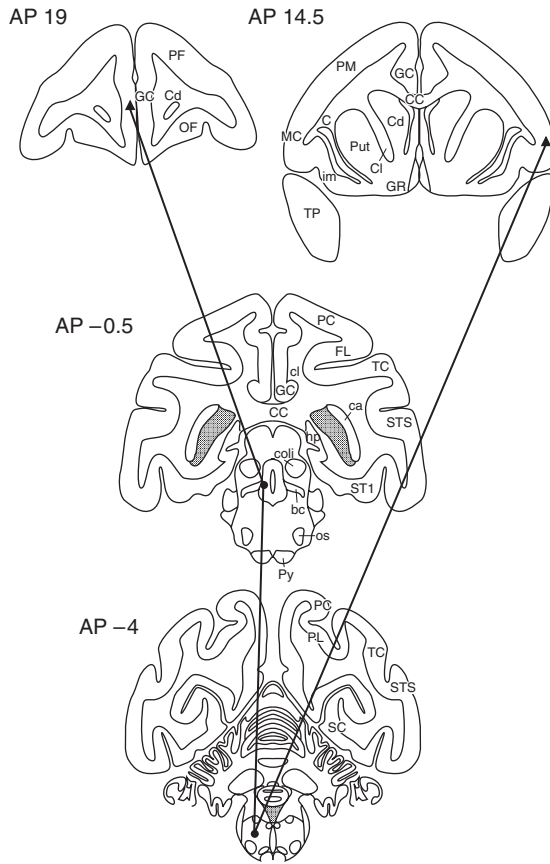


Figure 5. Scheme of a monkey brain as in Figure 2. Triangles indicate sites the electrical stimulation of which produces vocalisation (AP 19) or isolated vocal fold movements (AP 14.5), respectively. Square (AP -0.5) indicates injection site capable of blocking limbically but not neocortically induced vocal fold activity. Circle (AP -4) indicates site of laryngeal motoneurons. The connecting lines represent monosynaptic connections in the case of the projection from the anterior cingulate cortex to the peri-aqueductal grey, and polysynaptic connections to the other two cases (Jürgens & Zwirner, 1996).

system is closely tied to neocortical functions, as demonstrated in the two clinical cases.

The drastic effect of damage to the anterior cingulate area on human emotional speech is explained further by the fact that the anterior nucleus of the thalamus, which provides a large limbic source for the anterior cingulate area, has in humans substantially increased in volume over what would be expected in an ape thalamus of human brain size (Armstrong, 1986).

These brief remarks underline the indispensable participation of limbic functions in human speech. Limbic emotional vocal expressions and neocortically controlled speech work together in tandem. Only humans are able to act out their emotions with a species-specific motor system: the speech system. This explains why, for instance, even very aggressive acts are often expressed as verbal attacks and why verbal expressions of emotions often lead to relief from the respective emotional state. Talking to each other, the highest form of social communication, may cause pleasure, not talking to each other may cause discomfort. Although human social communication is dominated by the cortically guided language, the limbic structures are always involved, as clearly seen in all vocal, facial, postural and other expressions of emotion.

A HYPOTHESIS

Considering the strengthening and outgrowing of the corticopyramidal tract in the phylogeny of mammals, especially in primates, the hypothesis is advanced that the last step in the evolution of the phonatory system in the brain was the outgrowing and augmenting of the fine fibre portion of the pyramidal system that serves the direct and fast innervation of the larynx and tongue muscles via the nucleus retroambiguus and nucleus hypoglossus. For the more effective movement of the vocal cords and the tongue, more neurones of the pyramidal cell type in the increasingly larger neocortical face area may have been recruited. The direct and voluntary control of vocal behaviour allows the fractionation of species-specific vocal patterns such as cooing and babbling, transforming them into imitated and successively learnt articulated vocal gestures, i.e. words (Levelt, 1989; Ploog, 1990, 1995). In a way, this process is reflected in human ontogeny during language acquisition, where the child not only gradually gains control over its babbling, mastering increasingly more phonemes of its mother tongue, but also separating the speech motor system from the rest of the body motor system. Indicative of this stage in development are the conspicuous so-called 'associative movements' of the extremities, especially the hands, which accompany the articulatory movements (Noterdame *et al.*, 1988).

DISCUSSION

Comment: Patients who are mute and have limited cortical lesions make an effort to express themselves in other ways, for example by gesture.

Ploog: Yes, that is correct, but these people have lost their desire to communicate.

Comment: So then there is no lesion that leaves a person trapped in their brain trying to communicate but can't?

Ploog: They don't try to communicate. The drive to communicate is gone.

Comment: What did early hominids do? They wouldn't have used their hands for communication because they needed them for tool manufacture. You could turn that around and ask why tool manufacture was so static for so long and then it suddenly took off so quickly. You could reasonably argue that the reason it was so static for so long was that they needed their hands for communication.

Ploog: There must have been selective pressure for vocalisation. Why should it be since vocal communication among primates was so predominant and so forceful, why should we introduce gestures? Of course, when a chimp does a certain facial expression, it is also a kind of gesture, but it is an expression of emotion. This may mean that language and emotions are somehow related.

Comment: You mention how these primates communicate so well with lots of signals and do this on an interpersonal basis. This is also cortical and it is also fixed.

Ploog: It is not cortical; it is subcortical.

Comment: Signals that are elicited in fixed situations are not really like language.

Ploog: It is like language in that it serves as communication; but it is of course not language in that it is not grammar or syntax.

Comment: We laugh and grunt to communicate and these are not language and they are subcortical. Is this what you mean?

Ploog: Yes. Language is something different. Vocalisation evolved into speech, but it is not language. The argument cannot be turned around, however, to say that because language has syntax and grammar, it has nothing to do with vocalisation.

Comment: Yes, speech is obviously a form of language that has a lot to do with vocalisation.

References

- Armstrong, E. (1986) Enlarged limbic structures in the human brain: the anterior thalamus and medial mamillary body. *Brain Research*, **362**, 394–7.
- Chen, Y. & Jürgens, U. (1995) Phylogenetic trends in the projections of the cortical tongue area within primates. *European Journal of Neuroscience*, Supplement **8**, Abstract 53.04, 149.

- Cramon, D. von & Jürgens, U. (1983) The anterior cingulate cortex and the phonatory control in monkey and man. *Neuroscience and Biobehavioral Reviews*, **7**, 423–6.
- Hepp-Reymond, M.-C. (1988) Functional organization of motor cortex and its participation in voluntary movements. In: *Comparative Primate Biology*, Vol. 4: *Neurosciences* (eds H. D. Steklis & J. Erwin), pp. 501–624. New York: Alan R. Liss Inc.
- Herzog, M. & Hopf, S. (1984) Behavioral responses to species-specific warning calls in infant squirrel monkeys reared in social isolation. *American Journal of Primatology*, **7**, 99–106.
- Jürgens, U. (1976) Projections from the cortical larynx area in the squirrel monkey. *Experimental Brain Research*, **25**, 401–11.
- Jürgens, U. & Cramon, D. von (1982) On the role of the anterior cingulate cortex in phonation: a case report. *Brain and Language*, **15**, 234–48.
- Jürgens, U. & Lu, C.-L. (1993) The effects of periaqueductally injected transmitter antagonists on forebrain-elicited vocalization in the squirrel monkey. *European Journal of Neuroscience*, **5**, 735–41.
- Jürgens, U. & Ploog, D. (1970) Cerebral representation of vocalization in the squirrel monkey. *Experimental Brain Research*, **10**, 532–54.
- Jürgens, U. & Ploog, D. (1976) Ethologische Grundlagen. In: *Handbuch der Psychologie*, Vol. 8.1 (ed. L. J. Pongratz), pp. 599–633. Göttingen, Toronto, Zürich: Hogrefe.
- Jürgens, U. & Ploog, D. (1981) On the neural control of mammalian vocalization. *Trends in Neurosciences*, **4**, 135–7.
- Jürgens, U. & Pratt, R. (1979) The cingular vocalization pathway in the squirrel monkey. *Experimental Brain Research*, **34**, 499–510.
- Jürgens, U. & Zwirner, P. (1996) The role of the periaqueductal grey in limbic and neocortical vocal fold control. *Neuroreport*, **7**, 2921–3.
- Jürgens, U., Kirzinger, A. & Cramon, D. von (1982) The effects of deep-reaching lesions in the cortical face area on phonation. A combined case report and experimental monkey study. *Cortex*, **18**, 125–40.
- Kirzinger, A. & Jürgens, U. (1982) Cortical lesion effects and vocalization in the squirrel monkey. *Brain Research*, **233**, 299–315.
- Kuypers, H.G.J.M. (1958) Corticobulbar connexions to the pons and lower brain-stem in man. *Brain*, **81**, 364–88.
- Levelt, W.J.M. (1989) *Speaking: From Intention to Articulation*. Cambridge, MA and London: The MIT Press.
- Müller-Preuss, P. & Jürgens, U. (1976) Projections from the cingular vocalization area in the squirrel monkey. *Brain Research*, **103**, 29–43.
- Noterdame, M., Amorosa, H., Ploog, M. & Scheimann, G. (1988) Quantitative and qualitative aspects of associated movements in children with specific developmental speech and language disorders and in normal pre-school children. *Journal of Human Movements Studies*, **15**, 151–69.
- Phillips, C.G. (1979) The cortico-spinal pathway of primates. In: *Integration in the Nervous System* (eds H. Asanuma & V. J. Wilson), pp. 263–8. Tokyo: Igaku-Shoin.
- Ploog, D. (1981) Neurobiology of primate audio-vocal behavior. *Brain Research Reviews*, **3**, 61–76.
- Ploog, D. (1986) Biological foundations of the vocal expressions of emotions. In: *Emotion: Theory, Research and Experience*, Vol. III: *Biological Foundations of Emotion* (eds R. Plutchick & H. Kellerman), pp. 173–97. New York: Academic Press.

- Ploog, D. (1988) Neurobiology and pathology of subhuman vocal communication and human speech. In: *Primate Vocal Communication* (eds D. Todt, P. Goedecking & D. Symmes), pp. 195–212. Berlin, Heidelberg: Springer.
- Ploog, D. (1990) Neuroethological foundations of human speech. In: *From Neuron to Action* (eds L. Deecke, J. Eccles & V. Mountcastle), pp. 365–74. Berlin, Heidelberg, New York: Springer.
- Ploog, D. (1995) Neuroethological prerequisites for the evolution of speech. *Biology International*, Special Issue, **33**, 46–9.
- Ploog, D., Hupfer, K., Jürgens, U. & Newman, J.D. (1975) Neuroethologic studies of vocalization in squirrel monkeys with special reference to genetic differences of calling in two subspecies. In: *Growth and Development of the Brain* (ed. M.A.B. Brazier), pp. 231–54. New York: Raven Press.
- Sutton, D., Larson, C. & Lindeman, R.C. (1974) Neocortical and limbic lesion effects on primate phonation. *Brain Research*, **71**, 61–75.
- Winter, P., Ploog, D. & Latta, J. (1966) Vocal repertoire of the squirrel monkey (*Saimiri sciureus*), its analysis and significance. *Experimental Brain Research*, **1**, 359–84.
- Winter, P., Handley, P., Ploog, D. & Schott, D. (1973) Ontogeny of squirrel monkey calls under normal conditions and under acoustic isolation. *Behaviour*, **47**, 230–9.

