

Where did language come from? Precursor mechanisms in nonhuman primates

Josef P Rauschecker



At first glance, the monkey brain looks like a smaller version of the human brain. Indeed, the anatomical and functional architecture of the cortical auditory system in monkeys is very similar to that of humans, with dual pathways segregated into a ventral and a dorsal processing stream. Yet, monkeys do not speak. Repeated attempts to pin this inability on one particular cause have failed. A closer look at the necessary components of language, according to Darwin, reveals that all of them got a significant boost during evolution from nonhuman to human primates. The vocal-articulatory system, in particular, has developed into the most sophisticated of all human sensorimotor systems with about a dozen effectors that, in combination with each other, result in an auditory communication system like no other. This sensorimotor network possesses all the ingredients of an internal model system that permits the emergence of sequence processing, as required for phonology and syntax in modern languages.

Address

Department of Neuroscience, Georgetown University, Washington, DC 20057, USA

Corresponding author: Rauschecker, Josef P (rauschej@georgetown.edu)

Current Opinion in Behavioral Sciences 2018, 21:195–204

This review comes from a themed issue on **The evolution of language**

Edited by **Christopher Petkov** and **William Marslen-Wilson**

<https://doi.org/10.1016/j.cobeha.2018.06.003>

2352-1546/© 2018 Elsevier Ltd. All rights reserved.

Introduction

In 1866, the Linguistic Society of Paris banned any existing or future debates on the origin of language [1]. Thankfully, societies have changed since then, and we are permitted once again to ask where language comes from. With the advent of modern neurobiology, we can even ask the question in a much more pointed fashion than ever before and we do not dismiss it any longer as unanswerable, although it is still raising controversy.

Based on the dual-stream model of auditory cortical processing [[5[•]],3] (Figure 1), the present article will focus (1) on auditory word forms and how they are

mapped to meaning; (2) it will discuss how words are produced by combining and concatenating sequences of sounds in the appropriate order; and (3) how perception and production interact through an internal model system to enable communication and language at a higher order. The article does not attempt to cover the ultimate question of how thoughts are put into words ('Language is a system of signs that expresses ideas') [4] and how the rules for combining words ('syntax') are implemented in the brain. Although sequential analysis must be an essential mechanism for syntax as well [5[•]], actual precursor systems in nonhuman primates are harder to identify [6[•]]. This may be the consequence of a more limited auditory working memory span in monkeys [7], as sentence-level processing requires longer time windows and integration over greater distances.

Due to space constraints, the current piece clearly cannot do enough justice to the role of the basal ganglia and the cerebellum in the emergence of fine vocal motor control and sequence learning in humans. Their importance for the evolution of vocal learning and, thus, of language has been discussed elsewhere [8,9].

Neuroanatomical and functional components of auditory communication systems

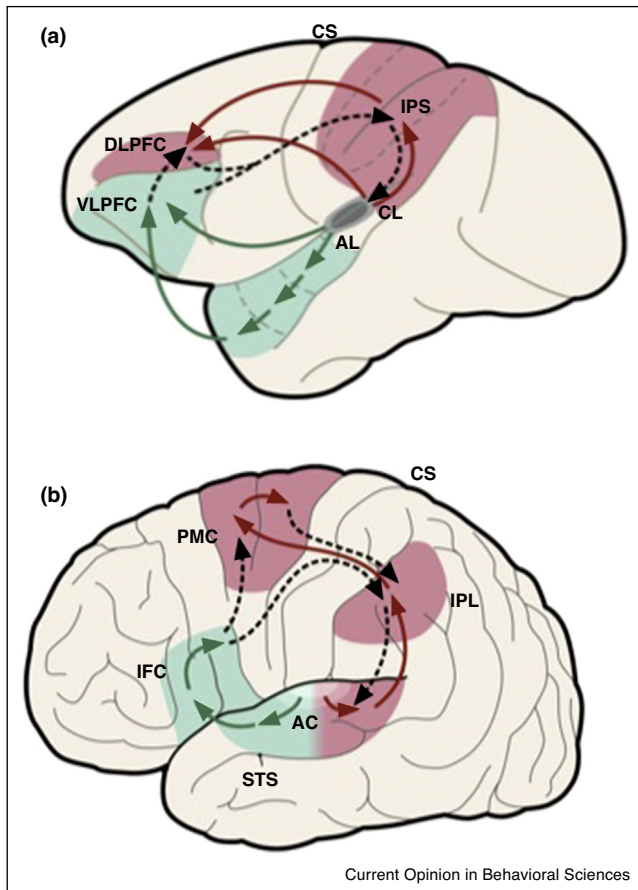
Most people would agree that human language as a complex cognitive system must have evolved from simpler systems that have existed in a common ancestor of present-day monkeys and humans. Although language as such does not exist in nonhuman primates, neurobiological precursor systems of language should, therefore, be identifiable. As already Darwin [10] pointed out, language consists of multiple discrete components (or subsystems) that emerged one at a time and, together, form the human language system. As a first step, it will be useful to compare the neuroanatomical components of auditory communication systems in monkeys and humans. Obviously, at a minimum, they have to consist of an auditory recipient structure and a vocal production system, which ideally should be tuned to each other, and a working memory system that can hold the communication signal in storage long enough to process sequences of sounds.

Auditory cortex: decoding of vocalizations and words

Functional organization of auditory regions in monkeys

After extensive preprocessing in the auditory periphery and brainstem the auditory communication signal arrives in the auditory cortex, where perception occurs. It was first shown in the rhesus monkey that early auditory

Figure 1



Dual-stream model of auditory processing. (a) Rhesus monkey [2,102–104]; (b) Human [3]. Abbreviations – CS: central sulcus; AL, CL: anterolateral, caudolateral belt; IPS: intraparietal sulcus; DLPFC, VLPFC: dorsolateral, ventrolateral prefrontal cortex.

cortex consists of a primary-like region (the ‘core’) and surrounding belt areas [11]. Although neurons in the core respond best to pure tones of a certain frequency, belt neurons prefer sounds of a greater bandwidth, such as band-passed noise bursts and frequency-modulated sweeps [12]. These sounds of intermediate complexity form the constituent elements of virtually all communication sounds. A certain percentage of neurons in the lateral belt of the rhesus monkey also respond to complete species-specific vocalizations, especially in the antero-lateral belt region (AL) [13]. Such neurons had previously been reported in the squirrel monkey [14,15].

Another region in the anterior part of superior temporal cortex of marmosets is tuned to the pitch of sounds and has, therefore, been conjectured to serve as a ‘pitch center’ [16]. Even more anterior to that location in the STG, a region was identified that prefers species-specific vocalizations over other vocalizations and sounds. This region not only showed sensitivity to the ‘voice’ of the

species, but also to the vocal identity of conspecific individuals [17].

Auditory subregions in humans

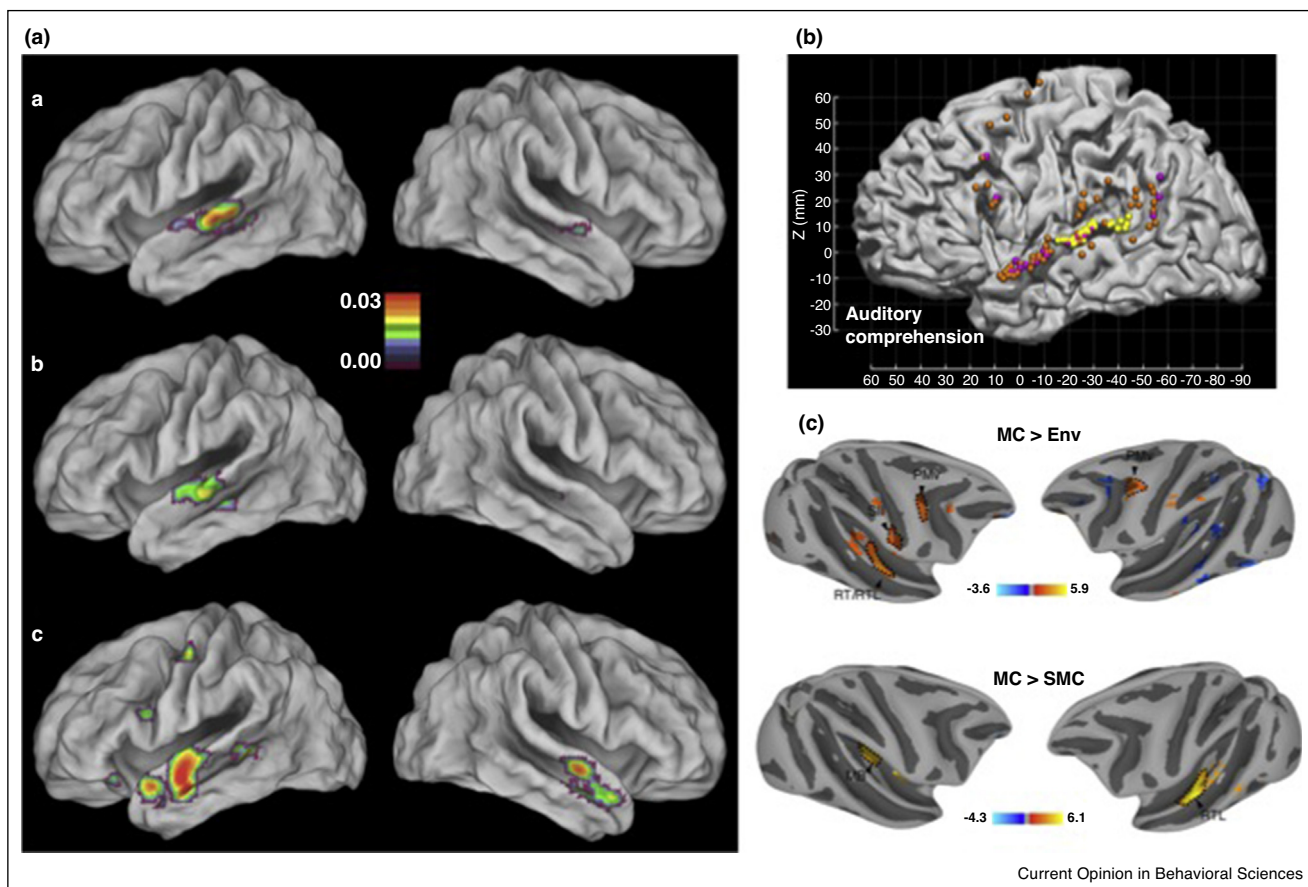
A very similar organization as in monkeys is found in human auditory cortex: A core region primarily responsive to tones is surrounded by a belt region tuned to BPN bursts [18], and an anterior parabelt region responds quite selectively to vowels and vowel-like sounds [19], consistent with earlier results [20]. The planum temporale near Heschl’s gyrus serves the role of a computational hub, distributing different types of features into the various subregions [21]. Higher regions of human auditory cortex along anterior superior temporal gyrus and sulcus (aSTG and aSTS) encode increasingly complex sounds, with specialized subregions responding best to musical instruments (on the right) and to speech sounds (on the left) [22,23*]. The parcellation of these multiple regions in functional terms, that is, their tonotopic organization [24*], follows closely a previously determined microanatomical parcellation in terms of cytoarchitectonics [25].

The voice region found in monkeys [17] (mentioned in the previous section) was first identified in humans using functional imaging [26]; responses include other non-linguistic vocalizations, such as weeping, crying and laughing. The existence of the pitch center found in monkeys was also verified in humans [27]. Taken together, this reveals an amazing similarity in the organization of higher, non-linguistic auditory cortex in monkeys and humans. Despite this recent imaging work and complementary electrocorticographic (ECoG) studies [28*], an accurate count of the number of auditory cortical areas in either monkeys or humans is still lacking. It cannot be excluded, therefore, that more areas have been added in humans that have further facilitated the evolution of speech perception. Finer tuning of auditory cortical neurons to critical speech sound features may also have contributed to an optimization of the system. In a more parsimonious view, however, monkey auditory cortex contains all the necessary ingredients for the decoding of speech sounds.

An ‘auditory word form area’ in the aSTG of humans? A gateway to semantics

Organization in terms of auditory linguistic information in humans also follows an antero-ventral trajectory: A meta-analysis of a large number of auditory imaging studies with PET and MRI found that phonemes are encoded antero-laterally from Heschl’s gyrus [29*]. Word representations are found anterior to that [29*,30,31] (Figure 2a). The resulting ‘word patch’ has occasionally been referred to as an auditory word form area [30] (AWFA), by analogy to the visual word form area (VWFA) for reading, which is found in ventral occipito-temporal cortex [32].

Figure 2



Converging evidence for the representation of vocalization sounds in the auditory ventral stream. **(a)** Humans, meta-analysis of auditory imaging studies using (a) phonemes, (b) words, (c) short phrases [29]; **(b)** Word representation in humans, identified by electrostimulation mapping of word comprehension [36]; **(c)** Functional MRI results of activation by species-specific vocalizations in the rhesus monkey [51]. *Abbreviations* – MC: monkey calls; ENV: environmental sounds; SMC: scrambled monkey calls; RTL: lateral rostrotemporal area.

The imaging data supporting an auditory word representation in the aSTG (see also [33,34] for earlier pioneering studies) have found independent confirmation by results of auditory single-unit and multi-unit responses to whole words in a human patient [35] and by electrostimulation mapping of word comprehension [36] (Figure 2b). The results also chime with voxel-based lesion-symptom mapping of stroke patients with semantic aphasia [37,38], and with theories based on studies of patients with primary progressive (semantic) dementia [39]. Both semantic patient groups show significant anterior superior temporal involvement, which is located even more anterior than the ‘word-form’ area [40]. Taken together, all of these data point to a hierarchically organized, anterior-directed system in the aSTG (extending laterally into the middle temporal region in humans [41]), which links auditory word representations with meaning [42,43,44**].

A seminal discovery more than three decades ago has shown that vervet monkeys also have the ability to map

complex sounds to meaning [45]: Alarm calls referring to different dangers result in differential behavior. The same is true for the monkeys’ own calls, which carry meaning about food types and social rank [46]. It will be interesting to ultimately compare the loci and extent of these auditory–semantic networks in both species.

The auditory ventral stream for auditory object recognition and categorization

The functional–anatomical pathway for encoding auditory object information, including short utterances, continues into ventrolateral prefrontal cortex (VLPFC) in monkeys [47] and ventral inferior frontal cortex (IFC) in humans [3], by way of the uncinate fasciculus and the extreme capsule fiber system (EmC). Together with the anterior part of the STG it has been referred to as the auditory ventral stream (AVS) in both species [[44**],3]. In monkeys, an antero-ventral network of vocalization-specific activation (without clear hemispheric asymmetry) is apparent in functional MRI studies [48–51] (Figure 2c).

There is good reason to believe that back-projections from frontal regions to aSTG in both species lead to categorization of sounds as auditory objects [44**].

The auditory dorsal stream for sensorimotor control and sequence analysis

As words are concatenated into longer sequences in humans and the rules of syntax have to be used to make sense out of whole sentences, speech turns into language and a process specific to humans comes into play. Most of this analysis is accomplished by the auditory dorsal stream (ADS) and its associated motor structures [52]. It is in this pathway where some authors have sought to find fundamental differences between monkeys and humans [53–55], but an auditory dorsal pathway with direct connections from posterior STG to prefrontal cortex clearly exists in both species [47,56]. It has been argued that even mere quantitative changes can lead to qualitative differences, as in nuclear physics or semiconductor electronics, once they surpass a certain threshold [57], but the search for the fundamental difference between monkeys and humans is still on, and concentrates now on the nature of the interaction between ventral and dorsal pathways [57–59].

Production of communication sounds and motor control of speech

Vocal communication in animals and humans, by definition, consists (a) of a vocal apparatus that is able to produce sounds, and (b) an auditory recipient structure in the brain that is able to decipher these sounds. The biological mechanics of sound-producing elements vary from species to species, thus limiting the possible sound structure of species-specific calls. By necessity, higher auditory structures in the brain of each species employing an auditory communication system should contain neurons that are able to respond to conspecific vocalizations or their elements, in order to understand them. Spoken language in humans uses acoustic signals that are produced by the vocal apparatus, consisting of the larynx and a large number of supralaryngeal articulators [60*], most notably the tongue. Besides these mechanical determinants, speech production is obviously determined by the neural control centers of the brain.

Motor control of vocal production: larynx and nucleus ambiguus

Much has been said about the apparent changes during evolution that may have occurred in the innervation of the larynx [9,61,62*,63]. In particular, it has been claimed that the neural control center of the larynx in the lower brainstem, the nucleus ambiguus, receives a direct projection from primary motor cortex M1 in humans but not in monkeys [61,63]. With recent advances in neuroanatomical tracing techniques [64] as well as diffusion tensor imaging [65], this has proven inaccurate, however: it turns out that at least one of the laryngeal muscles, the cricothyroid, which participates in the control of vocal pitch,

receives direct input from several motor cortical regions in the macaque, but its single most prominent cortical input comes from M1 [106]. In addition, diffusion tractography in both monkeys and humans shows that the laryngeal motor cortex (LMC) network is largely comparable in both species, albeit with more extensive LMC connections to somatosensory and parietal cortex in humans [65]. Thus, there again seems to be only a quantitative difference in LMC connections between monkeys and humans.

Furthermore, the nucleus ambiguus is only one of several brainstem nuclei controlling the >40 muscles of the vocal apparatus, and its function is restricted to the control of only 6 of them [62*]. As such, it does have some control over pitch changes and over the length of the vocal tract, as necessary for the production of different vowels [66] ('phonation'). Together with improved voluntary control of respiration and coordinated breathing, this may have led to a first simple system for volitional phonation in monkeys [67,68] and, perhaps, to the evolution of singing and yodeling in humans. Thus, speech and language may have in fact evolved from song and music, as some (including Darwin) have argued [69,70].

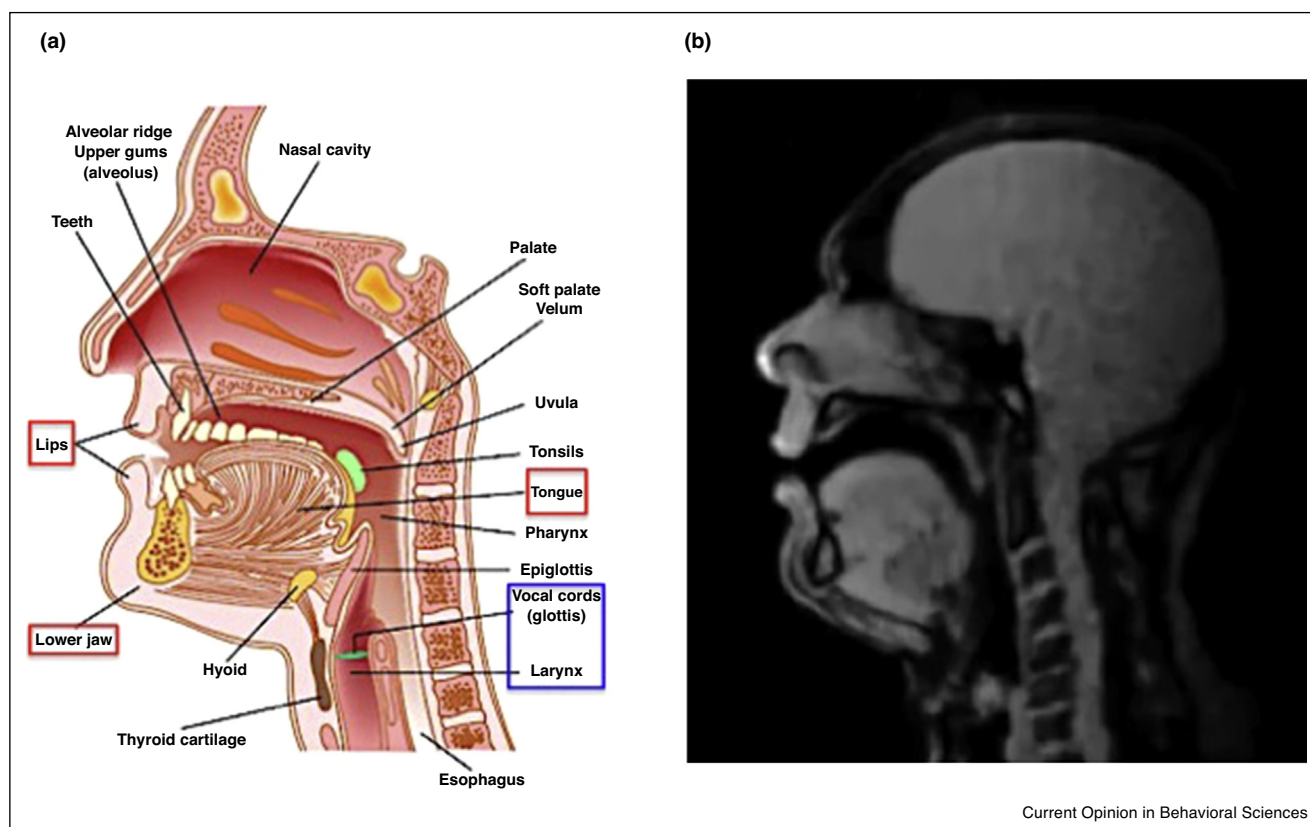
Supralaryngeal control of vocal production

Although the larynx is able to produce relatively simple vocalizations with varying pitch and spectral content, most of the information content of human speech is generated by the supralaryngeal components of the vocal tract, like tongue, lips, and lower jaw. These articulators are under the control of three cranial nerves (hypoglossal, facial and trigeminal) and their corresponding brainstem nuclei (Figure 3a). Surprisingly little is known about the cortical innervation of these nuclei in monkeys [71], but detailed studies with transneuronal tracers are currently under way (P Strick, personal communication). Initial results demonstrate converging input from two locations in primary motor cortex: from a ventral location in the orofacial region and a more dorsal location representing the diaphragm for breathing (Helou *et al.*: *Soc Neurosci*, 2017;406.03).

It was also found recently that posterior parietal cortex (PPC) provides surprisingly prominent input to the hand in rhesus monkeys, forming a higher-order 'command apparatus' for fine motor control of complex hand movements [72*], which might have enabled the evolution of tool use in primates. A similar command center in the PPC may have evolved for vocal control in humans but not in monkeys.

The supralaryngeal components of the human vocal tract enable the human vocal system to produce hundreds of discernable phonemes across the languages of the world and carry the bulk of the information content of human speech. Whereas the larynx itself can only produce

Figure 3



Current Opinion in Behavioral Sciences

Articulators of the human vocal apparatus. **(a)** Only larynx and vocal cords (framed by blue box) are innervated from the nucleus ambiguus (laryngeal nerve), whereas supralaryngeal articulators (framed by red boxes) are innervated from the hypoglossal (tongue), facial and trigeminal nerves (lips, mandible). **(b)** Mid-sagittal MRI slice through the human brain and skull. Real-time MRI permits the observation of articulator movements during song and speech [105].

Source: a: modified from MIT OpenCourseWare; b: from M Echternach, M Burdumi, L Traser, B Richter, Universitätsklinikum Freiburg, Germany (<https://www.youtube.com/watch?v=GCluRCd2YuM>).

temporal modulations of the voice through glottal closure and via pulsatile exhalation (cf. the barks, grunts and screams of macaques), the human vocal tract can produce subtle closures and constrictions of the airflow that each carry their own acoustic–phonetic signatures depending on place or manner of articulation (bilabial, dental, fricative, and so on). Combining the various phonemes leads to the combinatorial explosion that is language, and each language uses only a small subset of possible combinations.

The tongue carries much of the load for the production of these ‘information-bearing elements’ [73], and it is for good reason that many languages use the same word for ‘tongue’ and ‘language’ (lingua, langue). The supralaryngeal part of the vocal tract is the real reason for the enormous diversity of the vocal repertoire in humans and the multitude of speech sounds they can produce by combining coordinated movements of the various articulators [60]. The tongue alone (innervated by the n. hypoglossus) consists of eight individual muscles, which allow to move its tip in three spatial dimensions

and change its width and length as well as curl, protrude, and retract it (using the hyoid bone as a lever). Tongue action is supplemented by coordinated movements of the lips and the mandible, creating multiple points of contact with the gums and different points of closure of the airways responsible for the production of various types of sounds with distinct vocal signatures, including virtually all consonants. Consonants (in accord with their rarer occurrence) provide most of the information content in speech. In fact, a number of languages, such as Hebrew and Arabic, use consonantal alphabets (‘abjads’) to transcribe their spoken words with individual signs for consonants, but not for vowels.

Observation of speech movements in real time

Visualization of the intricate and speedy movements of the supralaryngeal elements of the vocal apparatus during singing and speaking has become possible through real-time magnetic resonance imaging [74,75] (rtMRI; Figure 3b; (<https://www.facebook.com/quartznews/>

[videos/1176008672432833/](#)). Watching these videos should convince anyone that speech is very likely the most sophisticated motor behavior in the human body, rivaled only by the movements of the hand and fingers, and undoubtedly many times more complex than in any other animal species that produces communication sounds. It appears highly unlikely, therefore, that monkeys, lacking most of the necessary articulators, could, in principle, produce human speech with their own vocal tracts, even if they had the neural wherewithal to do it [76]. Simple observation of this miraculous machine in action should be reason enough to at least consider the evolution of human speech production as one of the key events driving the evolution of speech and language.

Talking birds

Previous theories emphasizing the crucial importance of speech production for the evolution of language [77] have been dismissed by making the (perhaps facetious) argument that even parrots can learn to speak. The most publicized of such cases, an African grey parrot named Alex, was reported to have a vocabulary of 150 words, close to that of a 2-year old human infant [78]. Other bird species, such as mynahs and mockingbirds, which can also learn to imitate human speech sounds, have been reported to imitate artificial sounds, like car and machine sounds as well, so the primary purpose of these abilities is unlikely for communication.

Although parrots may be able to produce imitations of human words and utterances, these observations tell us more about human speech perception than about vocal learning in animals: Most imitations are acoustically highly degraded and can only be understood as speech by humans because their own recipient structure (the auditory cortex) is tuned to the expected targets and fills in the rest [79] (just as we can understand sinusoidal speech or speech sampled through a cochlear implant).

Imitation in human infants

How then does imitation work? How do we know what muscles to contract and which articulators to activate, in order to produce the speech sound we have in mind? The sounds stored in auditory cortex during early exposure to speech serve as targets to guide motor production. Auditory feedback is obviously necessary to assure that the produced copy at least approximates the original; the resulting error is then minimized over many iterations. In fact, it may take children several years to perfect their pronunciation. This error minimization is a critical part of the model [3,80]; it does not have to be conscious, but is more likely an implicit learning process. Once an acceptable transformation from perception to production is established by learning, the result is fixed (with an occasional refresher?) and becomes the ‘forward model’, which can be accessed quickly without sensory feedback (‘ballistically’, as it were).

Imitation also works for visual gestures and has given rise to the mirror neuron theory, in which neurons in the premotor cortex of monkeys are activated not only during a certain sensory-guided motor act, like reaching or grasping, but also when the monkey observes another monkey performing the same action [81]. Thus, the mirror system is able to predict the sensory consequences of actions, and this is not restricted to the visual domain.

Motor theories of speech perception

The most prominent version of a motor theory of speech perception was put forward by Liberman and colleagues (1967). What started out as an ingenious explanation for acoustic invariance in speech production by Liberman and by others [82], turned into a radical theory claiming “*perception of phonetic structure without translation from preliminary auditory impressions*” [83]. In its final instantiation [84], Liberman’s theory “. . . takes the speech elements to be articulatory gestures that are the primary objects of both production and perception. Those gestures form a natural class that serves a linguistic function and no other. Therefore, their **representations are immediately linguistic**, requiring no cognitive intervention to make them appropriate for use by the other components of the language system” [84].

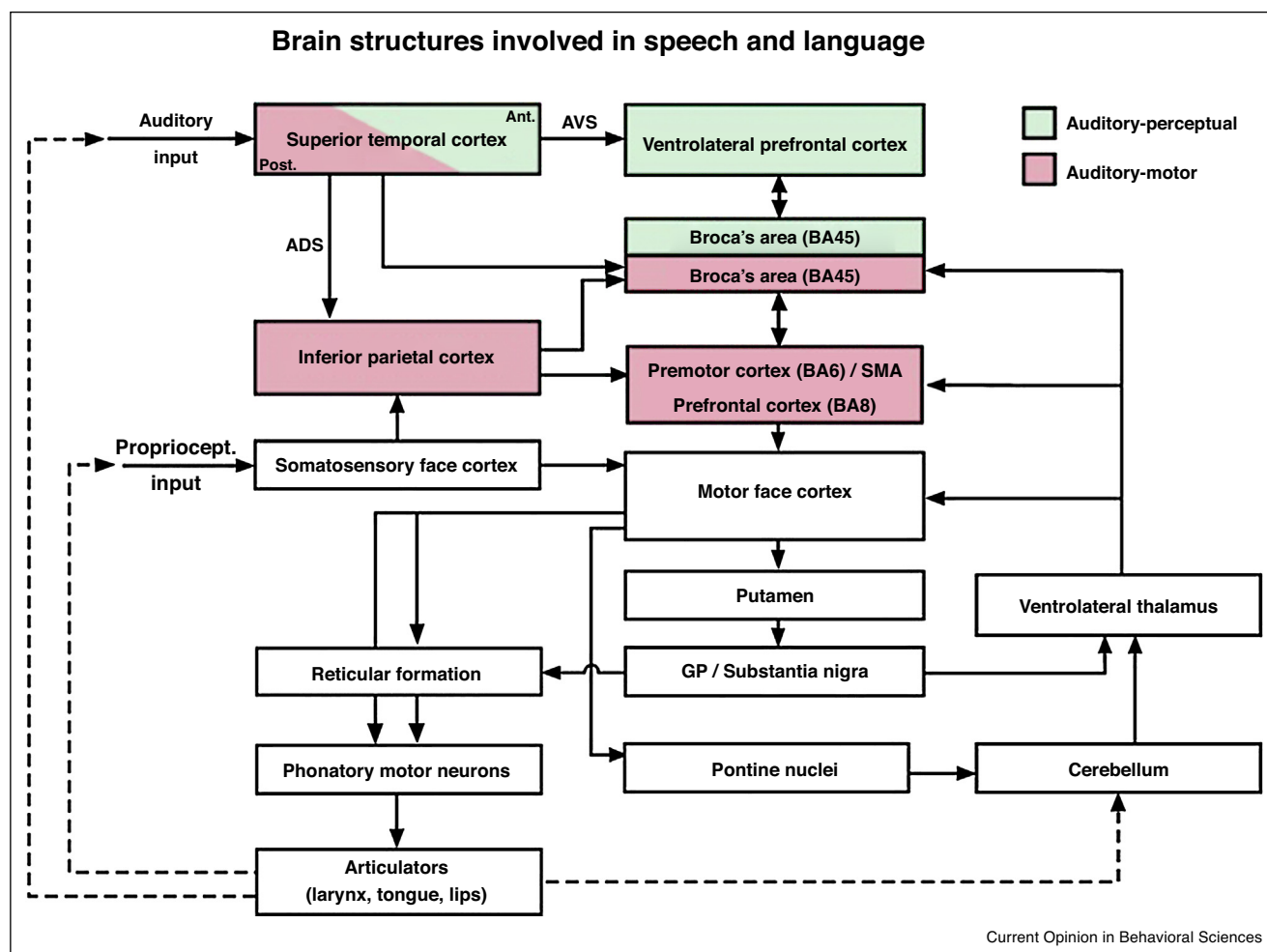
Does it follow then that we can understand only sounds that we can produce? More generally, are doable sounds represented differently in the brain? Is the ‘special’ aspect of speech the fact that doable sounds depend on sensorimotor (not just sensory) information? The control experiment would be to test sounds that cannot be imitated without extensive training, such as animal sounds [44**].

In the extreme, this would beg the question whether it is actually necessary that the motor system is intact for someone to learn and understand speech? Although it is well known that Broca’s area (as well as other motor-related regions) is active during speech perception [85,86], studies of Broca aphasics, who are unable to produce speech but can understand it, seem to argue against it [87]. However, more studies are needed, as the outcome may depend on the exact site of the lesion (BA44 or 45?) and on the level of speech processing. Clearly, the motor system is of great importance for the initial programming of internal models during language acquisition (as is an intact sense of hearing) [88,89*].

Convergence of ventral and dorsal streams

According to the dual-stream model of Rauschecker and Scott (2009) (Figures 1 and 4), the ventral and dorsal streams converge in Broca’s area in the inferior frontal cortex. The dorsal stream targets BA 44, while the ventral stream ends in BA 45. It has been conjectured that Broca’s region plays an important role in the transformation of highly processed auditory-sensory information into motor-articulatory signals and vice versa [80,90]. The dorsal stream in itself is organized as a system for sensorimotor integration and control and is programmed by sensory input

Figure 4



Brain structures involved in speech and language. Dual-stream model of Rauschecker and Scott [3] is superimposed on block diagram of speech production (modified and expanded from [62]). Auditory-perceptual structures (AVS) are shown in green, auditory-motor structures (ADS) in red. Motor structures, including basal ganglia, cerebellum, and brainstem nuclei, are displayed at the bottom as open blocks. Feedback connections are shown as dashed lines. *Abbreviations* – ADS: auditory dorsal stream; AVS: auditory ventral stream; BA: Brodmann area; SMA: supplementary motor area; GP: globus pallidus; Ant.: anterior; Post.: posterior.

from the ventral stream. The inferior parietal cortex measures the error signal and functions to minimize it.

Internal models

This transformation of sensory into motor information is typical for internal models and specifically forward models according to motor control theory [3,80,91–94]. According to Bornkessel-Schlesewsky *et al.* [52] the dorsal stream contains multiple internal models on different time scales: from articulation of single words to setting up sentence structures. As such it also plays a crucial role for sentence comprehension [58].

Working memory

This leads us to the final component which must have been essential for the evolution of language: the ability

to hold fleeting auditory information in short-term memory for as long as necessary. While this feature is already important for multisyllabic words in modern languages, it is even more essential for the speaking and understanding of long grammatical sentences. In fact, some authors have noted that monkeys have a significantly shorter memory span for sounds than for visual objects [7], and the anatomical substrate for auditory working memory in the medial temporal lobe may be reduced.

Conclusions and future studies

As we have seen, many of the structural components that make up the human language system are in place in nonhuman primates. Monkeys have a well-developed auditory cortex, which gives rise to two largely segregated

pathways: an antero-ventral and a postero-dorsal stream, which subserve different functions. The auditory ventral stream (AVS) works as a hierarchical network for the decoding of increasingly complex sounds. It is possible that the level of sound complexity and the tuning selectivity of auditory cortical neurons is higher in humans and the number of specialized regions within auditory cortex is higher than in monkeys, but this awaits further analysis measuring functional activity and connectivity in both species [95,96**]. The AVS then links sounds to meaning in networks of the anterior superior temporal cortex and temporal pole and engages top-down connections from prefrontal cortex for categorization of auditory objects [44**].

The auditory dorsal stream (ADS) is a complex network for sensorimotor integration and control, similar to its visual counterpart, in that it specializes for the analysis of space and motion, but also forms the basis for reaching and grasping in the visual domain and for audiomotor behavior, including speaking and singing in humans. As such, it encompasses aspects of the rival ‘where’, ‘how’ and ‘when’ models of the original dorsal stream [97*]. Although it undoubtedly exists in nonhuman primates [98] (and even in most other mammals), it seems as though the ADS has undergone massive expansion and refinement in humans that has enabled the control of a highly refined vocal apparatus, which has ultimately led to the production of speech.

For speech to evolve into language, a number of other things must have happened: The convergence of AVS and ADS in Broca’s area and the existence of longer memory spans, involving feedback loops from auditory to premotor regions, as well as feedback from somatosensory receptors informing internal models of the speech and vocal apparatus [99]. The ability to integrate observable visual speech articulations into auditory speech perception, most apparent when it goes wrong as in the McGurk effect, also resides in the dorsal stream [100,101]. More work on all these components will be necessary to come to a more comprehensive understanding of language and its evolution. Studies of language acquisition will take an especially important place. The picture that emerges is one of a slow/continuous converging evolution driven, as Darwin [10] surmised, by multiple factors, not by a single mutation that suddenly sparked the existence of language.

Conflict of interest statement

Nothing declared.

Acknowledgements

This work was supported by the National Institutes of Health [grant numbers R01NS052494, R01DC03489, and R01DC014989], by the National Science Foundation [grant numbers BCS-0519127, OISE-0730255], and by the Technische Universität München Institute for Advanced Study, funded by the German Excellence Initiative and the European Union Seventh Framework Programme under grant agreement n° 291763.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as

- of special interest
- of outstanding interest

1. Stam JH: *Inquiries Into the Origins of Language*. Harper and Row; 1976.
 2. Rauschecker JP, Tian B: **Mechanisms and streams for processing of “what” and “where” in auditory cortex**. *Proc Natl Acad Sci* 2000, **97**:11800-11806.
 3. Rauschecker JP, Scott SK: **Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing**. *Nat Neurosci* 2009, **12**:718-724.
 4. De Saussure F: **Course in general linguistics**. In *Course in General Linguistics (Cours de linguistique générale)*. Edited by Bally C, Sechehaye A. Fontana/Collins; 1916.
 5. Wilson B, Marslen-Wilson WD, Petkov CI: **Conserved sequence processing in primate frontal cortex**. *Trends Neurosci* 2017, **17**:72-81.
- Seminal nonhuman primate study on artificial grammar.
6. Dehaene S, Meyniel F, Wacongne C, Wang L, Pallier C: **The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees**. *Neuron* 2015, **88**:2-19.
- Excellent review on sequence processing.
7. Fritz J, Mishkin M, Saunders RC: **In search of an auditory engram**. *Proc Natl Acad Sci* 2005, **102**:9359-9364.
 8. Rauschecker JP: **Is there a tape recorder in your head? How the brain stores and retrieves musical melodies**. *Front Syst Neurosci* 2014, **8**.
 9. Petkov CI, Jarvis ED: **Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates**. *Front Evol Neurosci* 2012, **4**.
 10. Darwin CR: 1st edition. *The descent of man, and selection in relation to sex*. London: John Murray; 1871.
 11. Kaas JH, Hackett TA: **Subdivisions of auditory cortex and processing streams in primates**. *Proc Natl Acad Sci* 2000, **97**:11793-11799.
 12. Rauschecker J, Tian B, Hauser M: **Processing of complex sounds in the macaque nonprimary auditory cortex**. *Science* 1995, **268**:111-114.
 13. Tian B, Reser D, Durham A, Kustov A, Rauschecker JP: **Functional specialization in rhesus monkey auditory cortex**. *Science* 2001, **292**:290-293.
 14. Wollberg Z, Newman JD: **Auditory cortex of squirrel monkey: response patterns of single cells to species-specific vocalizations**. *Science* 1972, **175**:212-214.
 15. Winter P, Funkenstein HH: **The effect of species-specific vocalization on the discharge of auditory cortical cells in the awake squirrel monkey (*Saimiri sciureus*)**. *Exp Brain Res* 1973, **18**:489-504.
 16. Bendor D, Wang X: **The neuronal representation of pitch in primate auditory cortex**. *Nature* 2005, **436**:1161-1165.
 17. Petkov CI, Kayser C, Steudel T, Whittingstall K, Augath M, Logothetis NK: **A voice region in the monkey brain**. *Nat Neurosci* 2008, **11**:367-374.
 18. Wessinger CM, VanMeter J, Tian B, Van Lare J, Pekar J, Rauschecker JP: **Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging**. *J Cogn Neurosci* 2001, **13**:1-7.
 19. Chevillet MA, Riesenhuber M, Rauschecker JP: **Functional localization of the ventral auditory “what” stream hierarchy**. *J Neurosci* 2011, **31**:9345-9352.
 20. Obleser J, Boecker H, Drzezga A, Haslinger B, Hennenlotter A, Roetlinger M, Eulitz C, Rauschecker JP: **Vowel sound extraction**

- in anterior superior temporal cortex. *Hum Brain Mapp* 2006, **27**:562-571.
21. Griffiths TD, Warren JD: **The planum temporale as a computational hub.** *Trends Neurosci* 2002, **25**:348-353.
 22. Leaver AM, Rauschecker JP: **Cortical representation of natural complex sounds: effects of acoustic features and auditory object category.** *J Neurosci* 2010, **30**:7604-7612.
 23. Norman-Haignere S, Kanwisher NG, McDermott JH: **Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition.** *Neuron* 2015, **88**:1281-1296.
- Identification of distinct domains for the mapping of speech and music in human auditory cortex, replicating the earlier findings of [22] with updated fMRI techniques.
24. Leaver AM, Rauschecker JP: **Functional topography of human auditory cortex.** *J Neurosci* 2016, **36**:1416-1428.
- Complete fine-grained mapping of primary and nonprimary auditory cortical fields in human auditory cortex.
25. Morosan P, Rademacher J, Schleicher A, Amunts K, Schormann T, Zilles K: **Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system.** *Neuroimage* 2001, **13**:684-701.
 26. Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B: **Voice-selective areas in human auditory cortex.** *Nature* 2000, **403**:309-312.
 27. Norman-Haignere S, Kanwisher N, McDermott JH: **Cortical pitch regions in humans respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex.** *J Neurosci* 2013, **33**:19451-19469.
 28. Mesgarani N, Cheung C, Johnson K, Chang EF: **Phonetic feature encoding in human superior temporal gyrus.** *Science* 2014, **343**:1006-1010.
- Electrocorticographic mapping of phoneme features in epileptic patients.
29. DeWitt I, Rauschecker JP: **Phoneme and word recognition in the auditory ventral stream.** *Proc Natl Acad Sci* 2012, **109**:E505-514.
- Meta-analysis of functional imaging studies on the representation of phonemes and words in humans.
30. DeWitt I, Rauschecker JP: **Wernicke's area revisited: parallel streams and word processing.** *Brain Lang* 2013, **127**:181-191.
 31. DeWitt I, Rauschecker JP: **Convergent evidence for the causal involvement of anterior superior temporal gyrus in auditory single-word comprehension.** *Cortex* 2015, **77**:164-166.
 32. Cohen L, Dehaene S, Naccache L, Lehéricy S, Dehaene-Lambertz G, Hénaff MA, Michel F: **The visual word form area. Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients.** *Brain* 2000, **123**:291-307.
 33. Binder JR: **Human temporal lobe activation by speech and nonspeech sounds.** *Cereb Cortex* 2000, **10**:512-528.
 34. Scott SK, Blank CC, Rosen S, Wise RJS: **Identification of a pathway for intelligible speech in the left temporal lobe.** *Brain* 2000, **123**:2400-2406.
 35. Chan AM, Dykstra AR, Jayaram V, Leonard MK, Travis KE, Gygi B, Baker JM, Eskandar E, Hochberg LR, Halgren E *et al.*: **Speech-specific tuning of neurons in human superior temporal gyrus.** *Cereb Cortex* 2014, **24**:2679-2693.
 36. Roux FE, Miskin K, Durand JB, Sacko O, Réhault E, Tanova R, Démonet JF: **Electrostimulation mapping of comprehension of auditory and visual words.** *Cortex* 2015, **71**:398-408.
 37. Dronkers NF, Wilkins DP, Van Valin RD Jr, Redfern BB, Jaeger JJ: **Lesion analysis of the brain areas involved in language comprehension.** *Cognition* 2004, **92**:145-177.
 38. Shahid H, Sebastian R, Schnur TT, Hanayik T, Wright A, Tippett DC, Fridriksson J, Rorden C, Hillis AE: **Important considerations in lesion-symptom mapping: illustrations from studies of word comprehension.** *Hum Brain Mapp* 2017, **38**:2990-3000.
 39. Mesulam MM, Thompson CK, Weintraub S, Rogalski EJ: **The Wernicke conundrum and the anatomy of language comprehension in primary progressive aphasia.** *Brain* 2015, **138**:2423e7.
 40. Travis KE, Leonard MK, Chan AM, Torres C, Sizemore ML, Qu Z, Eskandar E, Dale AM, Elman JL, Cash SS *et al.*: **Independence of early speech processing from word meaning.** *Cereb Cortex* 2012 <http://dx.doi.org/10.1093/cercor/bhs228>.
 41. Davis MH, Johnsrude IS: **Hierarchical processing in spoken language comprehension.** *J Neurosci* 2003, **23**:3423-3431.
 42. Zhuang J, Tyler LK, Randall B, Stamatakis EA, Marslen-Wilson WD: **Optimally efficient neural systems for processing spoken language.** *Cereb Cortex* 2014, **24**:908-918.
 43. Kocagoncu E, Clarke A, Devereux BJ, Tyler LK: **Decoding the cortical dynamics of sound-meaning mapping.** *J Neurosci* 2017, **37**:1312-1319.
 44. Jiang X, Chevillet MA, Rauschecker JP, Riesenhuber M: **Training humans to categorize monkey calls: auditory feature- and category-selective neural tuning changes.** *Neuron* 2018, **98**:405-416.e4.
- Using two independent fMRI techniques, Jianget *al.* found that training to categorize novel monkey calls results in sharpened tuning to monkey-call features in left auditory cortex and category selectivity for different types of calls in lateral prefrontal cortex.
45. Seyfarth R, Cheney D, Marler P: **Monkey responses to three different alarm calls: evidence of predator classification and semantic communication.** *Science* 1980, **210**:801-803.
 46. Ghazanfar AA, Hauser MD: **The neuroethology of primate vocal communication: substrates for the evolution of speech.** *Trends Cogn Sci* 1999, **3**:377-384.
 47. Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP: **Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex.** *Nat Neurosci* 1999, **2**:1131-1136.
 48. Joly O, Pallier C, Ramus F, Pressnitzer D, Vanduffel W, Orban GA: **Processing of vocalizations in humans and monkeys: a comparative fMRI study.** *Neuroimage* 2012, **62**:1376-1389.
 49. Perrodin C, Kayser C, Logothetis NK, Petkov CI: **Auditory and visual modulation of temporal lobe neurons in voice-sensitive and association cortices.** *J Neurosci* 2014, **34**:2524-2537.
 50. Petkov CI, Kikuchi Y, Milne AE, Mishkin M, Rauschecker JP, Logothetis NK: **Different forms of effective connectivity in primate frontotemporal pathways.** *Nat Commun* 2015, **6**:6000.
 51. Ortiz-Rios M, Kuśmierk P, DeWitt I, Archakov D, Azevedo FAC, Sams M, Jääskeläinen IP, Keliris GA, Rauschecker JP: **Functional MRI of the vocalization-processing network in the macaque brain.** *Front Neurosci* 2015 <http://dx.doi.org/10.3389/fnins.2015.00113>.
 52. Bornkessel-Schlesewsky I, Schlesewsky M, Small SL, Rauschecker JP: **Neurobiological roots of language in primate audition: common computational properties.** *Trends Cogn Sci* 2015, **19**:142-150.
 53. Rilling JK, Glasser MF, Preuss TM, Ma X, Zhao T, Hu X, Behrens TEJ: **The evolution of the arcuate fasciculus revealed with comparative DTI.** *Nat Neurosci* 2008, **11**:426-428.
 54. Friederici AD: **Pathways to language: fiber tracts in the human brain.** *Trends Cogn Sci* 2009, **13**:175-181.
 55. Skeide MA, Friederici AD: **Response to Bornkessel-Schlesewsky *et al.* — towards a nonhuman primate model of language?** *Trends Cogn Sci* 2015, **19**:483.
 56. Frey S, Mackey S, Petrides M: **Cortico-cortical connections of areas 44 and 45B in the macaque monkey.** *Brain Lang* 2014, **131**:36-55.
 57. Bornkessel-Schlesewsky I, Schlesewsky M, Small SL, Rauschecker JP: **Response to Skeide and Friederici: the myth of the uniquely human "direct" dorsal pathway.** *Trends Cogn Sci* 2015, **19**:484-485.
 58. Griffiths JD, Marslen-Wilson WD, Stamatakis EA, Tyler LK: **Functional organization of the neural language system: dorsal and ventral pathways are critical for syntax.** *Cereb Cortex* 2013, **23**:139-147.
 59. Rolheiser T, Stamatakis EA, Tyler LK: **Dynamic processing in the human language system: synergy between the arcuate fascicle and extreme capsule.** *J Neurosci* 2011, **31**:16949-16957.

60. Levelt WJM: **Despite its publication year still the classic on all aspects of speaking..** *Speaking: From Intention to Articulation*. MIT Press; 1989.
61. Kuypers HGJM: **Corticobulbar connexions to the pons and lower brain-stem in man: an anatomical study.** *Brain* 1958, **81**:364-388.
62. Jürgens U: **Neural pathways underlying vocal control.** *Neurosci Biobehav Rev* 2002, **26**:235-258.
Classical review of neural pathways underlying vocal control.
63. Simonyan K: **The laryngeal motor cortex: its organization and connectivity.** *Curr Opin Neurobiol* 2014, **28**:15-21.
64. Rathelot J-A, Strick PL: **Subdivisions of primary motor cortex based on cortico-motoneuronal cells.** *Proc Natl Acad Sci* 2009, **106**:918-923.
65. Kumar V, Croxson PL, Simonyan K: **Structural organization of the laryngeal motor cortical network and its implication for evolution of speech production.** *J Neurosci* 2016, **36**:4170-4181.
66. Belyk M, Brown S: **The origins of the vocal brain in humans.** *Neurosci Biobehav Rev* 2017, **77**:177-193.
67. Hage SR, Nieder A: **Single neurons in monkey prefrontal cortex encode volitional initiation of vocalizations.** *Nat Commun* 2013, **4**.
68. Hage SR, Nieder A: **Dual neural network model for the evolution of speech and language.** *Trends Neurosci* 2016, **39**:813-829.
69. Fitch WT: *Musical Protolanguage: Darwin's Theory of Language Evolution Revisited*. 2009.
70. Fitch WT: **Empirical approaches to the study of language evolution.** *Psychon Bull Rev* 2017, **24**:3-33.
71. Morecraft RJ, Stilwell-Morecraft KS, Solon-Cline KM, Ge J, Darling WG: **Cortical innervation of the hypoglossal nucleus in the non-human primate (*Macaca mulatta*).** *J Comp Neurol* 2014, **522**:3456-3484.
72. Rathelot J-A, Dum RP, Strick PL: **Posterior parietal cortex contains a command apparatus for hand movements.** *Proc Natl Acad Sci* 2017, **114**:4255-4260.
Interesting study on projections of posterior parietal cortex to brainstem nuclei controlling the hand.
73. Suga N: **Philosophy and stimulus design for neuroethology of complex-sound processing.** *Philos Trans R Soc Lond B Biol Sci* 1992, **336**:423-428.
74. Niebergall A, Zhang S, Kunay E, Keydana G, Job M, Uecker M, Frahm J: **Real-time MRI of speaking at a resolution of 33 ms: undersampled radial FLASH with nonlinear inverse reconstruction.** *Magn Reson Med* 2013, **69**:477-485.
75. Lingala SG, Sutton BP, Miquel ME, Nayak KS: **Recommendations for real-time speech MRI.** *J Magn Reson Imaging* 2016, **43**:28-44.
76. Fitch WT, De Boer B, Mathur N, Ghazanfar AA: **Monkey vocal tracts are speech-ready.** *Sci Adv* 2016, **2**e1600723.
77. Liberman P: *The Biology and Evolution of Language*. Harvard University Press; 1984.
78. Pepperberg IM: **Animal language studies: what happened?** *Psychon Bull Rev* 2017, **24**:181-185.
79. Rauschecker JP, Shannon RV: **Sending sound to the brain.** *Science* 2002, **295**:1025-1029.
80. Rauschecker JP: **An expanded role for the dorsal auditory pathway in sensorimotor control and integration.** *Hear Res* 2011, **271**:16-25.
81. Rizzolatti G, Arbib MA: **Language within our grasp.** *Trends Neurosci* 1998, **21**:188-194.
82. Blumstein SE, Stevens KN: **Acoustic invariance in speech production: evidence from measurements of the spectral characteristics of stop consonants.** *J Acoust Soc Am* 1979, **66**:1001-1017.
83. Liberman AM, Mattingly IG: **The motor theory of speech perception revised.** *Cognition* 1985, **21**:1-36.
84. Liberman AM, Whalen DH: **On the relation of speech to language.** *Trends Cogn Sci* 2000, **4**:187-196.
85. Wilson SM, Saygin AP, Sereno MI, Iacoboni M: **Listening to speech activates motor areas involved in speech production.** *Nat Neurosci* 2004, **7**:701-702.
86. Pulvermuller F, Huss M, Kherif F, Moscoso del Prado Martin F, Hauk O, Shtyrov Y: **Motor cortex maps articulatory features of speech sounds.** *Proc Natl Acad Sci* 2006, **103**:7865-7870.
87. Hickok G, Costanzo M, Capasso R, Miceli G: **The role of Broca's area in speech perception: evidence from aphasia revisited.** *Brain Lang* 2011, **119**:214-220.
88. Bruderer AG, Danielson DK, Kandhadai P, Werker JF: **Sensorimotor influences on speech perception in infancy.** *Proc Natl Acad Sci* 2015, **112**:13531-13536.
89. Cheung C, Hamilton LS, Johnson K, Chang EF: **The auditory representation of speech sounds in human motor cortex.** *Elife* 2016, **5**e12577.
A paper summarizing some of the recent studies on electrocorticography (ECoG) of speech representations in humans.
90. Koehlin E, Jubault T: **Broca's area and the hierarchical organization of human behavior.** *Neuron* 2006, **50**:963-974.
91. Wolpert DM, Ghahramani Z, Jordan MI: **An internal model for sensorimotor integration.** *Science* 1995, **269**:1880-1882.
92. Franklin DW, Wolpert DM: **Computational mechanisms of sensorimotor control.** *Neuron* 2011, **72**:425-442.
93. Rauschecker JP: **Evolution of human language.** In *International Encyclopedia of Social & Behavioral Sciences*. Edited by Wright JD. Elsevier; 2015:323-328.
94. Rauschecker JP: **Auditory and visual cortex of primates: a comparison of two sensory systems.** *Eur J Neurosci* 2015, **41**:579-585.
95. Neubert F-X, Mars RB, Thomas AG, Sallet J, Rushworth MFS: **Comparison of human ventral frontal cortex areas for cognitive control and language with areas in monkey frontal cortex.** *Neuron* 2014, **81**:700-713.
96. Mars RB, Verhagen L, Gladwin TE, Neubert FX, Sallet J, Rushworth MFS: **Comparing brains by matching connectivity profiles.** *Neurosci Biobehav Rev* 2016, **60**:90-97.
Connectivity study using the same techniques in monkeys and humans.
97. Rauschecker JP: **Where, when, and how: are they all sensorimotor? Towards a unified view of the dorsal pathway in vision and audition.** *Cortex* 2018, **98**:262-268.
A review comparing anatomy and function of visual and auditory pathways in monkeys and humans.
98. Lomber SG, Malhotra S: **Double dissociation of "what" and "where" processing in auditory cortex.** *Nat Neurosci* 2008, **11**:609-616.
99. Tremblay S, Shiller DM, Ostry DJ: **Somatosensory basis of speech production.** *Nature* 2003, **423**:866-869.
100. Erickson LC, Zielinski BA, Zielinski JEV, Liu G, Turkeltaub PE, Leaver AM, Rauschecker JP: **Distinct cortical locations for integration of audiovisual speech and the McGurk effect.** *Front Psychol* 2014, **5**:534.
101. Erickson LC, Heeg E, Rauschecker JP, Turkeltaub PE: **An ALE meta-analysis on the audiovisual integration of speech signals.** *Hum Brain Mapp* 2014 <http://dx.doi.org/10.1002/hbm.22572>.
102. Rauschecker JP: **Processing of complex sounds in the auditory cortex of cat, monkey, and man.** *Acta Otolaryngol* 1997, **532**:34-38.
103. Rauschecker JP: **Cortical processing of complex sounds.** *Curr Opin Neurobiol* 1998, **8**:516-521.
104. Rauschecker JP: **Parallel processing in the auditory cortex of primates.** *Audiol Neurootol* 1998, **3**:86-103.
105. Echterhach M, Burk F, Burdumy M, Traser L, Richter B: **Dynamic changes of vocal tract articulators in different loudness conditions.** *PLoS One* 2016, **11**:e0153792.
106. Cerkevich CM, Strick PL: *Soc Neurosci* 2017, **408**:12.