Dual neural network model of speech and language evolution: new insights on flexibility of vocal production systems and involvement of frontal cortex

Steffen R Hage

Human speech vastly outperforms primate vocal behavior in scope and flexibility making the elucidation of speech evolution one of biology’s biggest challenges. A proposed dual-network model including a volitional articulatory motor network originating in the prefrontal cortex that is capable of cognitively controlling vocal output of a phylogenetically conserved primary vocal motor network attempts to bridge this gap. By comparing neuronal networks in human and non-human brains, crucial biological preadaptations are found in monkeys for the emergence of a speech system in humans. This model can explain behavioral evidence for vocal flexibility in cognitive tasks as well as during vocal development in monkeys as intermediate steps in the continuous evolution of speech in the primate lineage.

Address
Neurobiology of Vocal Communication, Werner Reichardt Centre for Integrative Neuroscience, University of Tübingen, Otfried-Müller-Str. 25, 72076 Tübingen, Germany

Corresponding author: Hage, Steffen R (steffen.hage@uni-tuebingen.de)

Introduction
The evolution of the human speech and language system is one of the most difficult and controversial biological questions, mainly because human speech considerably surpasses primate vocal communication systems in scope and flexibility with seemingly no counterpart in the animal kingdom, even among hominids [1,2,3]. Human vocal development is driven by learning, resulting in speech signals that are under volitional control and flexibly used [4]. At a very early stage, human infants start using largely innate, prelinguistic, speech-like vocalizations, so-called protophones, and non-speech-like vocalizations, such as crying and moaning, [5,6], which are then followed by vocalizations that consist of continuous or interrupted phonations, called babbling [7], that becomes increasingly speech-like during the first year [5]. In contrast, vocalizations of non-human primates are largely innate and stereotypic. Consequently, non-human primates have been largely overlooked as a model system for human speech in the last few decades. However, even if human speech production is more complex than the vocal motor system of non-human primates [2,8], evolutionary theories postulate pre-adaptations in the primate lineage, no matter how exiguous they might be [2]. In fact, several behavioral and neurophysiological studies in the last few years indicate that such inevitable preadaptations are present for vocal motor control mechanisms in non-human primates indicating continuous phylogenetic trends in the evolution of human speech in the primate lineage [1,2,9].

In the present article, I will first review recent behavioral, anatomical, and physiological insights suggesting that monkeys possess the rudiments to cognitively control their vocal output, an obligatory precursor for speech and language production [2]. I will also introduce a recently proposed dual-network model of speech and language evolution, which consists of executive control structures within the frontal lobe that are capable of taking control over ancient vocal pattern-generating and limbic networks [2]. Second, I will examine recent behavioral evidence suggesting experience-dependent acoustic changes of vocal behavior during development as an essential preadaptation for speech acquisition in human speech [5] and suggest a possible role of the dual network model in this matter.

Cognitive flexibility in primate vocal production as a preadaptation for human speech
While human speech is above all a learned vocal pattern, the vocal repertoire of our closest relatives, non-human primates, consists mainly of stereotyped and largely innate calls that are uttered affectively in most contexts [2,3,8,9]. These assumptions were primarily supported by previous studies showing that monkeys that were deaf-born or deafened [10], raised in social isolation [10], or cross-fostered [11] developed the full adult species-specific call repertoire with only minor differences in call structure from the vocal utterances of their normally raised conspecifics. Furthermore, several studies found no differences in vocal pattern production in monkeys
after lesioning brain regions homologous to brain structures that are crucial for speech pattern production in humans [12–14]. Taken together, these studies indicate that monkeys do not learn their vocal utterances due to vocal imitation but rather suggests a largely innate structure of vocal utterances [3,15].

However, several recent studies indicate that monkeys have the ability to volitionally control when, but not how, to produce distinct vocalizations in a specific cognitive, social, or environmental context. This type of vocal behavior, also known as call usage learning [15,16], is deemed a critical preadaptation for the development of a flexible communicative system in the primate lineage [1,9]. This type of vocal learning is not dependent on learning new call patterns and enables the ability to withhold or initiate a specific vocalization, although it is still tied to the respective (motivational) context [17,18], or the more elaborate ability to decouple calls from the accompanying motivational state such that the subjects are able to use the calls in a novel context [2,9,16]. Several behavioral studies report that monkeys can volitionally initiate vocal output and instrumentize their calls in a goal-directed manner. For example, monkeys can vocalize or remain silent when exposed to operant conditioning tasks [19–21,22]. These findings support behavioral studies that show monkeys can produce or withhold alarm calls depending on the social context [17], avoid calling during masking external acoustic events [23,24], and show great flexibility within specific vocal communication situations [25]. In recent studies, Hage and colleagues demonstrated that rhesus monkeys are capable of selectively emitting different call types in response to distinct visual cues [26,27]. At first, they showed that monkeys can be trained in a visual detection task to vocalize whenever a colored visual cue appeared on the screen, showing consistent vocal performance. In addition, this study revealed that rhesus monkeys can be trained to switch between two distinct vocalizations (‘coos’ and ‘grunts’) on command on a trial-by-trial basis in a visual discrimination task. These results indicate that monkeys have rudimentary control over their vocal repertoire and, therefore, that rhesus monkeys can decouple their innate calls in a goal-directed way from the corresponding state of arousal to perform a specific task successfully. Furthermore, several studies observed volitional changes of vocal parameters such as vocal duration, amplitude, and frequency [22,26]. However, in none of the cases where operant conditioning was employed to assess vocal flexibility did the animals produce a vocal pattern de novo. Instead, they shifted the median distribution of distinct call parameters within the natural boundaries of their innate repertoire.

In summary, these studies indicate that monkeys possess preadaptations, such as call usage learning, that are crucial for the evolution of a learned vocal communication system, such as human speech, but lack the ability to learn or imitate new vocal signals [9,16,28]. Furthermore, these findings suggest a cognitive neuronal network that is capable of taking control over a basic vocal motor network that is producing largely innate vocal utterances in rather affective contexts.

### Dual neural network model underlying vocal motor control in monkeys and humans

Recently, a dual neural network model was suggested that postulates two structurally and functionally distinct parts [2], a volitional articulatory motor network (VAMN) originating in the prefrontal cortex (PFC) that cognitively controls the vocal output of a phylogenetically conserved primary vocal motor network (PVMN) mainly situated in subcortical structures (Figure 1). This dual network model suggests the linking of prefrontal and premotor structures with the vocalization system as a key neurobiological event and preadaptation for the evolution of speech and language in the primate lineage [2]. This hypothesis is supported by several comparative anatomical and physiological studies demonstrating that the basic architectonic plans of the ventrolateral PFC (vlPFC) are similar in both human and non-human primates despite considerable development of the human vlPFC. All previous findings of behavioral preadaptations within the primate lineage, such as volitional control of vocal onset or the capability for rudimentary modifications in call structure in non-human primates, might be explained by phylogenetic adaptations and modifications within the proposed dual neural network model.

First, all primates possess a PV MN that produces species-specific vocalizations with a largely fixed structure in non-human primates and non-verbal vocalizations in humans. This PV MN consists of two structurally and functionally distinct parts: a highly conserved vocal pattern-generating system in the brainstem, which coordinates all muscles involved in vocal production and an upstream limbic vocal-initiating network driving the pattern generator based on affective states (Figure 1) [2,8,9,29]. It is important to realize that the phylogenetically conserved primary vocal motor network is still involved in vocalization in humans [2,9,30]. Several studies revealed that one of its functions is to produce non-verbal vocal utterances such as crying, laughing, or moaning, all of which are largely innate and affective vocalizations considered to be directly homologous to monkey vocalizations [9,31]. In addition, the PV MN is also important for speech. It is active during voice speech production [32] and lesions within parts of PV MN, such as the anterior cingulate cortex or the periaqueductal gray, lead to severe deficits in speech production from monotonous intonation of speech signals to the point of mutism [33,34].

Second, a cortical network including several structures in the vlPFC, as well as the premotor and/or primary motor...
cortices is involved in taking cognitive control over the PVMN (Figure 1) [19**,35]. Here, the above-mentioned conditioning experiments where monkeys were trained to vocalize on command allowed these neuronal preadaptations of cognitive control of vocal behavior in the monkey brain to be investigated. Recently, several studies have suggested a putative homologue of Broca’s area in the vlPFC of rhesus monkeys (areas 44 and 45) that might play a leading role in the cognitive control of vocal behavior. Electrical microstimulation in monkey area 44 elicited orofacial responses [36], and auditory stimulation with species-specific vocalizations evoked distinct patterns of brain activity in monkey area 44 and 45 in awake behaving rhesus monkeys [37,38] that were comparable to responses in the human Broca’s area [39]. In humans, Broca’s area is one of the crucial structures involved in speech production [40]. Using direct cortical recordings, recent studies have revealed that Broca’s area in humans is predominantly activated before the utterance of a speech signal but is silent during the corresponding articulation (Figure 2). In contrast, increases in neuronal activity started right before speech onset and lasted throughout vocal production in the motor cortex [41*]. Similar functional patterns seem to already exist in the monkey brain. Vocalization-related activity that specifically predicts the preparation of instructed vocalizations in response to a visual cue were recorded in the putative homologue of the human Broca’s area [19**] and ventral premotor cortex [19**,20]. As in the human brain, areas 44 and 45 showed neuronal activity predominantly before vocal output and not while the monkeys were vocalizing. In contrast, neurons in the premotor cortex changed their neuronal activity directly before vocal onset and lasted throughout vocal production (Figure 2). These findings indicate that Broca’s area is involved in the coordination of speech initiation rather than in producing speech output directly and suggest a cardinal role of the monkey homologue of human Broca’s area in vocal planning and call initiation, which are putative phylogenetic preadaptations in the primate
lineage for such speech control mechanisms in humans. Interestingly, recent studies have found neurons in the prefrontal and premotor cortex of freely moving and communicating marmoset monkeys that were active prior and during some vocal utterances [42,43]. These findings are another compelling indicator that monkeys de facto use volitional control of vocal output spontaneously and not only during operant conditioning tasks.

**Impact of auditory feedback on cognitive vocal motor control: new insights from primate vocal development**

For several decades, primate vocalization has been thought to be largely predetermined, undergoing little-to-no experience-dependent acoustic changes during vocal development (see above). If changes occurred, they were solely (and quite reasonably) attributed to growth or...
sometimes also increased motor control [44,45]. Therefore, it was concluded that the development of acoustic call structure does not require any auditory experience in monkeys [9,28]. Recent studies, however, which focused on changes in specific call parameters, such as call entropy (which corresponds to the noisiness of the sound), which are known from other vertebrate species to undergo distinct changes during vocal development [46,47], revealed the role of auditory feedback in vocal development in marmoset monkeys (Figure 3a). This New World monkey species typically lives in family groups, where mothers typically give birth to twins, and both the father and older siblings carry the young [48]. When separated, both young and adult subjects emit vocalizations that play
a significant role in re-establishing contact. Contact calls by one party typically elicit counter-calling from the other party, which constitutes an excellent opportunity to study the effects of parental feedback on the vocal development of the young. During ontogeny, marmoset vocalizations undergo distinct changes in several call parameters, such as decreases in call frequency and entropy and increases in call duration and inter-call intervals [49–51]. Contingent parental auditory feedback accelerates the vocal development of mature vocalizations [48]. A recent study revealed that siblings that have been experimentally exposed to more contingent parental vocal feedback in the first two postnatal months transformed their contact calls faster from noisy, immature calls (with high entropy) to tonal adult-like calls (with low entropy) than their twin sibling experiencing less contingent feedback [52**] (Figure 3b). Another recent study investigated whether such parental auditory feedback is an obligate requirement for proper vocal development or whether it simply accelerates vocal development, without a detrimental effect if absent, by comparing the vocal behavior of two sets of offspring. One set of infants was normally raised, while the other was separated from the parents after the third postnatal month [53*]. All the monkeys eventually produced mature vocalizations. In contrast to normally raised monkeys, however, marmosets with limited parental feedback also produced infant-specific vocal behavior, which is characterized by specific call types, distributions and transitions (Figure 3c). Taken together, these findings indicate that marmoset monkeys require contingent parental auditory feedback for both vocal usage learning, that is, learning which call types to use or not to use in specific behavioral contexts, as well as production-related vocal learning.

The underlying neural network that is involved in vocal usage and production-related learning in non-human primates is yet unknown. Studies in vocal learners such as birds and humans show that interconnections between auditory and vocal-motor side are crucial for song and speech acquisition, respectively. In humans, strong interconnections crucial for speech production and acquisition exist between the frontal lobe articulatory network and temporal lobe auditory structures [54]. Very recently, a direct connection between the avian analog of the mammalian premotor cortex (HVC) and the avian analog of the secondary auditory cortex (Av in the caudal mesopallium) has also been identified in song birds as being essential for vocal imitation learning [55*]. In monkeys, connections between auditory cortical regions and prefrontal and premotor structures have also been identified [56]. In addition, both of these cortical areas include single neurons exhibiting audio–vocal interaction, that is, showing auditory response properties combined with a vocalization-related activity preceding vocal output [57,58]. Therefore, an anatomically plausible loop that is capable of modulating vocal pattern generating mechanisms during development might project from the auditory cortex to ventrolateral prefrontal cortex to premotor cortex to pontine reticular formation, encompassing the vocal pattern generating network [29], with all areas serving as potential hubs in audio–vocal modulation [41*,57,58]. These findings suggest essential preadaptations for the direct manipulation of vocal motor output via the VAMN as may be required during development. In addition, it will be of great interest if and how cortico-basal loops, which are crucial for speech and song acquisition in vocal learners such as humans and song birds [9,59], are capable of modulating the VAMN, for example, via the prefrontal and/or premotor cortex, to permanently change vocal pattern generating mechanisms in the PVMN. Further studies in developing monkeys will now have to elucidate the causal role of such audio–vocal networks and cortico-basal loops in experience-dependent modifications of the acoustic call structure during vocal development in non-human primates, a putative phylogenetic preadaptation in the primates for speech acquisition in humans.

**Conclusion**

The emerging link between prefrontal executive structures and the ancient vocal motor system might be one of the key preadaptations for the evolution of human speech in the primate lineage. On the basis of comparative neurophysiological and anatomical studies, the proposed dual-network model reveals that distinct neuronal preadaptations essential for human speech evolution are also present in monkeys. In addition, the model can also explain all observed behavioral preadaptations in non-human primates, such as cognitive vocal motor control and the presence of changes in acoustic call structure during vocal development, two crucial precursors in the continuous evolution of human speech in the primate lineage. In future, non-human primates will be a valuable model to better understand such preadaptations in monkeys that are essential for human speech evolution and to decipher common neuronal networks involved in both monkey vocalization and human speech.

**Conflict of interest statement**

This work was supported by the Werner Reichardt Centre for Integrative Neuroscience (CIN) at the Eberhard Karls University of Tübingen (CIN is an Excellence Cluster funded by the Deutsche Forschungsgemeinschaft within the frame-work of the Excellence Initiative EXC 307) and by the Deutsche Forschungsgemeinschaft (Grant HA5400/3-1). The funders had no role in review design, decision to publish, or preparation of the manuscript.

**Acknowledgements**

I thank Adeen Flinker for providing high resolution images that were used in Figure 2, Asif A Ghazanfar for providing high resolution images that were used in Figure 3 and John Holmes for proofreading.
References and recommended reading

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

• of special interest
•• of outstanding interest


A detailed review of behavioral, anatomical, and neurophysiological findings suggesting that monkeys possess the rudiments to cognitively control their vocal output resulting in a dual-network model able to develop a neurobiological scaffold for speech evolution in primates.


This is the first study recording single-unit activity in the prefrontal cortex of vocalizing monkeys. Call-related neurons specifically predicted the preparation of cognitively controlled vocal output, but not from spontaneous calls.


41. Flinker A, Korzeniewska A, Shestyuk AV, Franaszczuk PJ, Dronkers NF, Knight RT, Crone NE: Redefining the role of
Broca’s area in speech. Proc Natl Acad Sci U S A 2015, 112:2871-2875.

This study revisits the classic notion of the role of Broca’s area in human speech. Using direct cortical surface recordings in neurosurgical patients, the authors provide evidence that Broca’s area is involved in coordinating the transformation of information across large-scale cortical networks involved in spoken word production rather than (as previously thought) the production of speech signal itself which seems to be implemented by motor cortex.


This study presents the first experimental evidence for production-related vocal learning during development in monkeys. Marmoset infants experiencing more contingent auditory feedback had a faster rate of vocal development than their siblings.


This study observed that a lack of parental feedback after the third postnatal month preserves infant-like vocal behavior beyond infancy in marmoset monkeys.


This study provides neurophysiological and neuroanatomical evidence for a cortical motor to auditory circuit being essential for vocal imitation learning and the adaptive modification of vocal timing in a song bird.


