Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Current Opinion in Behavioral Sciences

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)

Current Opinion in Behavioral Sciences 2018, **21**:80–87

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.2018.02.010

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 instrumentalize their calls in a goal-directed manner. For example, monkeys can vocalize or remain silent when exposed to operant conditioning tasks [19••,20–22]. 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 base 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 PVMN that produces species-specific vocalizations with a largely fixed structure in non-human primates and non-verbal vocalizations in humans. This PVMN 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 PVMN is also important for speech. It is active during voiced speech production [32] and lesions within parts of PVMN, 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