The origins of the vocal brain in humans

Michel Belyk[1,2], Steven Brown[1]

[1]Department of Psychology, Neuroscience & Behaviour, McMaster University, Hamilton, Ontario, Canada
[2]Department of Neuropsychology & Psychopharmacology, Maastricht University, Maastricht, Limburg, The Netherlands

RUNNING TITLE: Origins of vocal brain

Correspondence to:
Steven Brown, Ph.D.
Department of Psychology, Neuroscience & Behaviour
McMaster University
1280 Main St. West
Hamilton, ON, L8S 4K1
Canada
Phone: 1 (905) 525 9140 extension 23892
e-mail: stebro@mcmaster.ca

The evolution of vocal communication in humans required the emergence of not only voluntary control of the vocal apparatus and a flexible vocal repertoire, but the capacity for vocal learning. All of these capacities are lacking in non-human primates, suggesting that the vocal brain underwent significant modifications during human evolution. We review research spanning from early neurophysiological descriptions of great apes to the state of the art in human neuroimaging on the neural organization of the larynx motor cortex, the major regulator of vocalization for both speech and song in humans. We describe changes to the location, structure, function, and connectivity of the larynx motor cortex in humans compared with non-human primates, including critical gaps in the current understanding of the brain systems mediating vocal control and vocal learning. We explore a number of models of the origins of the vocal brain that incorporate findings from comparative neuroscience, and conclude by presenting a summary of contemporary hypotheses that can guide future research.

Keywords: vocalization, brain, evolution, larynx motor cortex, vocal learning, human, primate

Vocal communication in humans is characterized by a number of distinct features not found in non-human primates or most other animals, including voluntary control of vocal behavior, the acquisition of vocal repertoires through imitative learning, parallel channels of vocal communication through speech and song, the generation of phonological structure through combinatorial mechanisms, and cultural transmission of vocal information, among many others (Arbib, 2012; Christiner & Reiterer, 2013; Gracco & Löfqvist, 1994; Kuhl & Meltzoff, 1996; Merker, Morley, & Zuidema, 2015; Patel, 2003, 2008). A critical question for human evolution is how these capacities emerged. We will focus here on phylogenetic changes to the structure and function of the human brain, with an emphasis on the neural mechanisms of vocalization for both speech and song. In particular, we will examine the evolutionary changes that have occurred to the *larynx motor cortex* (LMC), which is the primary cortical center for vocalization in the human brain (Bouchard, Mesgarani, Johnson, & Chang, 2013; Breshears, Molinaro, & Chang, 2015; Brown et al., 2009; Brown, Ngan, & Liotti, 2008; Loucks, Poletto, Simonyan, Reynolds, & Ludlow, 2007; Simonyan, Ostuni, Ludlow, & Horwitz, 2009; reviewed in Conant, Bouchard, & Chang, 2014; Simonyan, 2014; Simonyan & Horwitz, 2011). We will deal with two major evolutionary issues, first how the larynx motor cortex of humans evolved from a non-vocal LMC precursor, and second how humans acquired the capacity for vocal learning from an ancestral species that lacked this capacity. While these two changes can be thought of as independent evolutionary events (Ackermann, Hage, & Ziegler, 2014), we will discuss novel models that attempt to establish an evolutionary connection between cortical control of vocalization and the capacity for vocal learning.

1. Anatomy and physiology of the larynx

The larynx is the organ of vocalization in mammals, with innervation coming from the branchiomotor division of the vagus nerve (Jürgens, 2002). The larynx is composed of four principal cartilages, one bone, a set of intrinsic laryngeal muscles that interconnect them, and various extrinsic laryngeal muscles that connect them with the rest of the skeleton. Figure 1 depicts some of the relevant anatomy of the larynx.

Insert Figure 1 about here

Suspended within the larynx are the vocal folds. A principal role of the larynx across animal species is to serve as a form of airway protection. Forceful compression of the vocal folds forms a secondary closure of the airway below that of the epiglottis (Ardran & Kemp, 1952). Another major function of the vocal folds is to act as the primary sound-source for vocal communication. The vocal folds are composed of the body of the thyroarytenoid muscle and vocal ligament, enveloped in a membranous covering. The body and cover together form a non-linear dynamic system that vibrates in a complex and periodic fashion when air passes through the space between the two vocal folds, a space known as the glottis (Story & Titze, 1995; Titze & Story, 2002). This process of sound production through vocal-fold vibration is referred to variously as vocalization, phonation, and voicing. The resultant complex waves are filtered as they pass through the oral cavity by the action of articulators such as the lips and tongue in order to produce the diverse array of sounds that compose speech (Fant, 1960; Titze, 2008).

There are three major dimensions of movement within the larynx (Seikel, King, & Drumwright, 2010). First, the glottis can be opened or closed by separating the vocal folds (abduction) or bringing them together at the midline (adduction), respectively. Whereas passive breathing requires an open glottis and thus vocal fold abduction, vocalization requires adduction as an initial step, so as to bring the vocal folds into the air stream and allow them to be set into vibration by expiratory airflow. Contraction of the posterior cricoarytenoid muscle abducts the vocal folds by pivoting the horn-shaped arytenoid cartilages. Contraction of the lateral cricoarytenoid reverses this action, and contraction of the interarytenoid muscles draws the paired arytenoid cartilages towards the midline, both effecting vocal fold adduction (Gray, 1918).

Second, starting from an adducted vocalization-ready position, the vocal folds can either be stretched, causing them to vibrate at a higher fundamental frequency (F0), or they can be relaxed, causing them to vibrate at a lower F0. Stretching and relaxing the vocal folds affects the frequency at which these membranes vibrate by altering certain physical properties, such as their stiffness, thickness, and tension, among others (Titze & Story, 2002). These physical parameters are controlled primarily by the cricothyroid (CT) and thyroarytenoid (TA) muscles. Contraction of the CT muscle rocks the thyroid cartilage forward, thereby stretching and increasing the tension of the vocal folds, and causing them to vibrate at a higher frequency (Buchthal, 1959; Gay, Hirose, Strome, & Sawashima, 1972; Hollien & Moore, 1960; Kempster, Larson, & Kistler, 1988; Roubeau, Chevrie-Muller, & Saint Guily, 1997). The TA muscle may relax the vocal folds, and in that sense acts as an antagonist to the CT muscle to decrease F0. However, part of the TA muscle lies within the body of the vibrating mass of the vocal folds such that contraction of the TA muscle may increase vocal-fold stiffness, and increase F0. The net influence of the TA muscle in either increasing or decreasing F0 strongly depends on non-linear interactions with the CT muscle, the range of frequencies being produced, vocal register, and expiratory force

(Kochis-Jennings, Finnegan, Hoffman, Jaiswal, & Hull, 2014; Lowell & Story, 2006; Titze, Luschei, & Hirano, 1989).

Most vocal communication in humans relies on the these two dimensions of vocal-fold movement, namely the rapid cycling between adducted and abducted positions for the alternation between voiced and unvoiced sounds, and the stretching or relaxing of the vocal folds to determine vocal pitch. However, the larynx can also be moved vertically within the vocal tract by the action of the extrinsic laryngeal muscles. Two sets of muscles pull the larynx in roughly opposing directions along the vertical axis. Laryngeal elevators raise the larynx during swallowing and vomiting so as to close the airway (Ardran & Kemp, 1952; Lang, Dana, Medda, & Shaker, 2002). These muscles extend from the larynx to more-superior structures (Gray, 1918; Seikel et al., 2010), including the mandible, pharynx, tongue, and temporal bone. Laryngeal depressors, also known as strap muscles, lower the larynx during yawning (Barbizet, 1958). These muscles extend from the larynx to more-inferior structures (Seikel et al., 2010), including the sternum and scapula. Untrained singers raise or lower the larynx as they modulate vocal pitch (Pabst & Sundberg, 1993; Roubeau et al., 1997), although these movements have only a modest influence on F0 (Sapir, Campbell, & Larson, 1981; Shipp & Izdebski, 1975; Vilkman, Sonninen, Hurme, & Körkkö, 1996). Vertical laryngeal movements may have a more prominent effect on the apparent size of a speaker’s vocal tract, which is a cue to his/her body size (Pisanski et al., 2014).

2. Larynx motor cortex

2.1 A brief history of the search for the human larynx motor cortex

Voluntary control of the larynx is mediated by the primary motor cortex in the precentral gyrus of the frontal lobe, which gives rise to a descending corticobulbar projection to the nucleus ambiguus in the medulla, which itself sends out motor neurons to the skeletal muscles of the larynx via the myelinated portion of the vagus nerve that has been implicated in social communication more broadly (Porges, 2001). The location of the larynx-controlling region of the motor cortex was controversial for much of the 20th century. Foerster (1931) observed that electrical stimulation of the subcentral gyrus (and adjacent Rolandic operculum) elicited a grunting or groaning sound, although he did not report eliciting more speech-like vocalizations. Classic work by Penfield and Boldrey (1937) in analyzing the homunculus of the human primary motor cortex through neurosurgical stimulation of the brain of awake patients did not directly observe the laryngeal muscles during these procedures and so did not localize a specific larynx-controlling region. This was due to the fact that Penfield did not record from the intrinsic laryngeal muscles during his procedures, as well as because he was not able to identify a specific location for vocalization compared to related oral functions. While he was able to elicit rudimentary vocalizations in some of his patients, this invariably occurred in combination with movement of other orofacial effectors, such as the lips and/or tongue. Hence, Penfield assigned vocalization to a large swath of the orofacial motor cortex, rather than to a unique location in the way that he had done for the other effectors of the body.

Our understanding of LMC localization changed in the 21st century with the first functional magnetic resonance imaging (fMRI) studies looking specifically at laryngeal functioning. Starting in the early 1990's, brain imaging research began to describe the networks involved in speaking and singing (see Turkeltaub, Eden, Jones, & Zeffiro, 2002, for an early meta-analysis). However, speech and song are highly complex sequences of movements, involving rapid and coordinated movements of the respiratory and articulatory musculature, in addition to the larynx. The early neuroimaging studies made no distinction between the phonatory and articulatory components of speech.

Interest in identifying a larynx-specific motor cortical representation emerged using the combination of transcranial magnetic stimulation (TMS) and electromyography (EMG), which was developed for the neurological assessment of cranial nerve function (Thumfart, Pototschnig, Zorowka, & Eckel, 1992). In contrast to the earlier neurosurgical studies, which used vocalization as a proxy for laryngeal-muscle stimulation, TMS/EMG studies were able to directly measure physiological responses in the intrinsic laryngeal muscles. TMS permitted the elicitation of motor responses in two of the intrinsic laryngeal muscles that contribute directly to the control of vocal pitch, namely the CT muscle and the TA muscle. The scalp locations where stimulation had its maximum effect were 7.5 ± 1.4 cm and 10.3 ± 1.9 cm along the interaural-plane for the CT and TA muscles, respectively (Rödel et al., 2004). However, the more dorsal location of the TA muscle overlapped with the location of the tongue (10.5 ± 0.8 cm) from a separate experiment reported by the same group (Rödel, Laskawi, & Markus, 2003).

In an fMRI experiment, Loucks et al. (2007) observed that vocalization engaged the same areas of motor cortex as silent expiration, in locations consistent with an earlier positron emission tomography study of respiration (Ramsay et al., 1993; see also Simonyan et al., 2009; Kryshtopava et al., 2017). This finding suggested that the motor control of the laryngeal muscles is highly integrated with the driving force for vocalization, namely expiration. This linkage between vocalization and expiration (but not inspiration) in the human motor cortex is consistent with the observation that oral sound production in humans has evolved to occur almost exclusively on expiration (i.e., it is egressive), with ingressive sound production being relatively rare (e.g., gasping). This is in contrast to vocalizing in many primate species that occurs biphasically on both inspiration and expiration (Geissmann, 2000). In fact, MacLarnon & Hewitt (1999) observed that the thoracic vertebral column of humans – which contains spinal motor circuits mainly for expiration – is allometrically enlarged in humans compared to hominins and modern-day primates, which they argued was an adaptive change for respiratory control, including for vocalization (see also MacLarnon & Hewitt, 2004).

Brown, Ngan & Liotti (2008) performed an fMRI study that attempted to identify a specific somatotopic location for the larynx in the human motor cortex distinct from the representation of the articulatory muscles, not least in light of the uncertainties of Penfield's neurosurgical findings and the apparent overlap of the larynx and articulators in the later TMS studies. In particular, they carried out a direct comparison between vocalization and non-vocal laryngeal movements (i.e., forceful adduction of the vocal folds via glottal stops) in the same participants. As a somatotopic reference, they also had participants perform lip and tongue movement, since Penfield obtained much more reliable localizations for these effectors. Importantly, glottal stops and vocalization led to strongly overlapping activations in a region of primary motor cortex that Loucks et al. (2007) had previously identified as integrating vocal and expiratory functions, leading them to dub the common area of activation as the "larynx/phonation area". This region was found to be directly adjacent to the somatotopic lip area in the dorsal part of the orofacial motor cortex. In other words, the area for phonation was found to be close to, but distinct from, an area for the control of articulation.

Belyk & Brown (2014) later found that this same region contained a representation of not only the intrinsic musculature of the larynx but also the extrinsic musculature that moves the entire larynx vertically within the airway, although more-ventral regions of the motor cortex made a stronger contribution to such vertical movement. This observation in humans is similar to the representation of the extrinsic laryngeal muscles near the LMC of monkeys (Hast, Fischer, & Wetzel, 1974). The larynx motor cortex thus controls the three principal dimensions of laryngeal movement.

Overall, it appears that evolutionary reorganization of the human motor cortex has brought the three major components of vocalization – namely expiration, phonation, and articulation – into close proximity, perhaps creating what some theorists refer to as a "small-world architecture" (Sporns, 2006; Sporns & Zwi, 2004), whereby networks function most efficiently when they have an abundance of short-distance or local connections, supplemented with relatively few long distance connections.

Bringing this field full circle to the surgical studies of Penfield in the 1930's, more-recent neurosurgical research has replicated Penfield's original finding that vocalization can be elicited through stimulation of the human primary motor cortex in a location similar to that observed in the brain imaging studies of vocalization (Breshears et al., 2015). Likewise, it was observed that a more-ventral location was active in anticipation of the onset of vocalization or during changes to ongoing vocal patterns (Chang, Niziolek, Knight, Nagarajan, & Houde, 2013). A neurosurgical study that recorded local field potentials in the brains of awake patients during syllable production produced an important finding: the human precentral gyrus contains *not one but two representations of the laryngeal muscles*, both of which are distinct from the adjacent articulatory muscles (Bouchard et al., 2013). A similar duality has been observed on the basis of gene expression profiles in postmortem human brains (Pfenning et al., 2014). The more dorsal of the larynx areas observed by Bouchard et al. (2013) was located in the dorsal part of the orofacial primary motor cortex, concordant with the later findings of Breshears et al. (2015) as well as the prior neuroimaging studies of laryngeal functioning (Brown et al., 2008; Grabski et al., 2012; Loucks et al., 2007; Olthoff, Baudewig, Kruse, & Dechent, 2008; Peck et al., 2009; Simonyan et al., 2009). The second larynx area was found at the ventral extreme of the orofacial motor cortex, in the subcentral gyrus and Rolandic operculum, concordant with the original findings of Foerster (1931).

The LMC of humans, thus, appears to have a two-part structure that is not present in other primates. This novel duality of the human LMC raises important questions about the relative functional roles of the LMCs in controlling the laryngeal muscles across biological functions (e.g., airway protection and vocal functions). In light of this dual representation of the larynx within the motor cortex, we will follow the terminology of Pfenning et al. (2014) in referring to these areas as the dorsal and ventral LMCs, respectively, as shown schematically in Figure 2.

Insert Figure 2 about here

Intriguingly, the ventral LMC in the Rolandic operculum corresponds to the location that was predicted to be the location of the human LMC from a comparative neuroscience perspective (Ludlow, 2005), even before neuroimaging studies addressed this question empirically. The ventral LMC is more proximate than is the dorsal LMC to the LMC of Old World monkeys (Simonyan & Jürgens, 2002), New World monkeys (Hast et al., 1974; Hast & Milojkvic, 1966; Jürgens, 1974), and great apes (Leyton & Sherrington, 1917). We will argue below in the section “Comparative neuroscience of the larynx motor cortex” that the ventral LMC is the human homologue of the non-human primate LMC. It should be noted that, although the monkey LMC is found in the premotor cortex (area 6v), rather than primary motor cortex (area 4), we will adopt the common label of "larynx motor cortex" across species in order to facilitate comparison (Simonyan & Jürgens, 2002, 2003).

2.2 Somatosensory cortex

The somatotopy of primary motor cortex in the precentral gyrus is paralleled by a similar, posteriorly-positioned map in the primary somatosensory cortex of the postcentral gyrus (Penfield & Boldrey, 1937; Penfield & Rasmussen, 1950). It is therefore likely that the dorsal and ventral LMCs are accompanied by dorsal and ventral larynx *sensory* areas (LSCs), although there has been considerably less research on laryngeal sensory representations in humans. Brain imaging studies on professional singers have described a dorsal LSC in the postcentral gyrus, directly posterior to the dorsal LMC. This area shows experience-dependent plasticity, with increased singing-related activation and decreased grey matter concentration in professional opera singers compared to novices (Kleber, Veit, Birbaumer, Gruzelier, & Lotze, 2010; Kleber et al., 2016).

While the central sulcus provides a clear anatomical landmark dividing the primary motor cortex from the primary sensory cortex, this separation is more ambiguous in the subcentral gyrus and Rolandic operculum, where cytoarchitectonic boundaries are more difficult to assess from gross anatomical landmarks. This anatomical region contains the borders of the primary motor cortex (BA 4) and primary somatosensory cortex (BA's 3/1/2), in addition to a distinct cytoarchitectonic zone of its own (BA 43). Based on his cytoarchitectonic observations, Brodmann (1909) remarked that BA 43 most resembled the cortex of the postcentral gyrus (i.e., primary somatosensory cortex). Vogt, however, who was Brodmann's collaborator and contemporary, classified the same region as most resembling the ventral precentral gyrus (i.e., orofacial primary motor cortex) based on his observations of myeloarchitecture (Judaš & Cepanec, 2010; Vogt, 1910).

The contentious status of the cyto- and myeloarchitecture of the subcentral gyrus and Rolandic operculum extends to the neurophysiology of this region. Using magnetoencephelography, Miyaji et al. (2014) observed somatosensory activations along much of the extent of the subcentral gyrus in response to a puff of air applied to the dorsal surface of the larynx. In contrast, this same area has been shown to have the defining neurophysiological feature of motor cortex: electrical stimulation of this region elicits a vocal motor response (Foerster, 1931; Penfield & Boldrey, 1937). Although no study has examined both the laryngeal motor and sensory functions of the subcentral gyrus in the same brain, one electrocorticography study observed both auditory and speech-motor responses under electrodes in the subcentral gyrus (Cogan et al., 2014). It remains to be determined whether these studies have described a single functional zone with both motor and sensory properties, or distinct motor and sensory zones, reflecting parallel motor and sensory somatotopic maps. In discussing laryngeal motor control in this article, we will continue to follow Pfenning et al. (2014) in using the term ventral LMC when referring to the region of the subcentral gyrus and Rolandic operculum that is associated with vocal-motor output.

2.3 Connectivity

Little research has been conducted to assess the connectivity of either the dorsal or ventral LMC regions in humans. Simonyan et al. (2009) carried out an analysis of both structural and functional connectivity of the dorsal LMC. Using diffusion tensor imaging (DTI), they observed that the dorsal LMC had structural connections principally within the ipsilateral orofacial primary motor and sensory cortices, with evidence of additional structural connectivity beyond the precentral and postcentral gyri being observed in only a minority of participants. Functional connectivity analyses revealed a more extensive and bilateral connectome. The dorsal LMC was positively associated with a network of speech-motor brain areas, including the inferior frontal gyrus (IFG), premotor cortex, auditory association cortex, supplementary motor area (SMA), pre-SMA, inferior parietal lobule (IPL), basal ganglia, and thalamus. Interestingly, the dorsal LMC was negatively associated with other cortical areas of specific importance for laryngeal motor control, including the ACC and a region of ventral M1. The discrepancy between structural and functional connectivity profiles may stem from the limitations of DTI tractography that render certain pathways difficult to detect. Alternatively, since functional-connectivity measures are sensitive to indirect relationships between brain areas, this network may include brain areas that are several synapses removed from the LMC.

A follow-up analysis (Kumar, Croxson, & Simonyan, 2016) estimated the location of the dorsal LMC from a meta-analysis, and observed a more extensive profile of structural connectivity that more closely matched the previously-observed pattern of functional connectivity. The authors further noted that, compared with the connectome of the monkey LMC, the human dorsal LMC has greater connectivity with parietal areas, including the somatosensory cortex and IPL. A meta-analysis from that study did not detect a ventral LMC location for tractography, which is consistent with the uncertainty about the existence of the ventral LMC from the published neuroimaging literature. One possible reason why few brain imaging studies report peak activations in the Rolandic operculum may be the tendency for strong activations from auditory cortex to blur across the Sylvian fissure, making it difficult to disentangle auditory from motor responses. In light of the recent neurosurgical observations of a dual structure of the LMC, it will be important to examine the possibility of differential patterns of structural connectivity between the ventral and dorsal divisions of the LMC, as well as potential connectivity between these two divisions within the precentral gyrus.

2.3.1 Inter-hemispheric connectivity

The larynx is a midline structure, and the two vocal folds operate as a coordinated pair to produce symmetrical and synchronous movements; asymmetrical movements of the vocal folds are indicative of pathology (Isshiki, Tanabe, Ishizaka, & Broad, 1977; Steinecke & Herzel, 1995). This symmetry is most likely supported by the bilateral innervation of the nucleus ambiguus by the LMC (Kuypers, 1958ab; Simonyan & Jürgens, 2003), although some stimulation studies have reported contralateral innervation of the intrinsic laryngeal muscles by both the motor cortex (Leyton & Sherrington, 1917) and nucleus ambiguus (Prades et al., 2012). However, this raises the question of whether the left and right LMCs differ in function and how they communicate with one another.

Inter-hemispheric fibers in division III of the corpus callosum connect the primary motor cortices of the two hemispheres (Fling, Benson, & Seidler, 2013; Hofer & Frahm, 2006). The inter-hemispheric fibers that link the left and right primary motor cortex (M1) are organized according to a motor homunculus similar to that in M1 itself, with the legs represented posteriorly in the corpus callosum and the face represented anteriorly (Wahl et al., 2007). Although the left and right LMC are connected via the corpus callosum in monkeys (Jürgens, 1976; Simonyan & Jürgens, 2002), human research has only demonstrated functional connectivity, but not structural connectivity, between the left and right dorsal LMC (Kumar et al., 2016; Simonyan et al., 2009), leaving the existence of inter-hemispheric fibers for either of the LMC regions in humans uncertain.

The axons of the motor corpus callosum for the upper limbs carry net inhibitory signals that are believed to facilitate the independent movement of the limbs on the two sides of the body (Netz, Ziemann, & Hömberg, 1995), so important for praxis. The utility of a mechanism supporting movement asymmetry is unclear for motor areas like the LMC that control the left and right vocal folds in a symmetrical and synchronous manner. Instead, it may be necessary to explore the hypothesis that LMC inter-hemispheric connections operate on principles that promote movement symmetry, rather than asymmetry.

Despite this need for symmetric activation of the two vocal folds, the innervation of most of the intrinsic laryngeal muscles by one of the branches of the vagus nerve is astoundingly asymmetric. In particular, the recurrent laryngeal nerve descends far below the level of the larynx to wrap around the lowest aortic arches, before ascending back up to innervate the intrinsic laryngeal muscles. Since the aortic arches themselves are asymmetrical, the path of this nerve is nearly twice as long on the left side as it is on the right (Prades et al., 2012). This would result in a drastic asynchrony in the timing of innervation of the two vocal folds were it not for a compensatory difference in the thickness of the nerves that helps offset this length difference (Krmpotic, 1959, cited in Walker, 1994). As a result of the balance between thickness and length, action potentials arrive at the laryngeal muscles with only a 2-4 millisecond difference between the left and right sides (Prades et al., 2012; Thumfart, 1988; Thumfart et al., 1992). Hence, despite an unusually asymmetric innervation pattern, the larynx motor system seems to operate in a bilateral fashion to innervate the two vocal folds in a symmetric and synchronous manner (Walker, 1994).

2.3.2 Corticobulbar connectivity

Of the known efferent connections of the LMC, the corticobulbar projection to the nucleus ambiguus, which itself contains the lower motor neurons that innervate the laryngeal muscles, has received the most attention. In monkeys, the LMC makes an indirect projection to the nucleus ambiguus via synapses in the reticular formation (Jürgens & Ehrenreich, 2007; Simonyan & Jürgens, 2003), while great apes have a sparse monosynaptic pathway from the LMC to the nucleus ambiguus (Kuypers, 1958a). This direct pathway is enlarged in humans, although it is still sparse relative to corticobulbar projections to other cranial-nerve motor nuclei (Iwatsubo, Kuzuhara, & Kanemitsu, 1990; Kuypers, 1958b). However, it is not known whether this direct projection originates from the dorsal or ventral LMC, since this distinction was not recognized at the time of these anatomical studies. While opportunities to study human connectivity using postmortem material are relatively rare, the efferent connectivity of the dorsal vs. ventral LMCs in the brain stem requires further investigation.

The evolutionary emergence of a direct pathway from the LMC to the nucleus ambiguus has been hypothesized by many to support efficient voluntary control of vocalization (Fischer & Hammerschmidt, 2011; Fitch, 2010, 2011; Fitch, Huber, & Bugnyar, 2010; Jarvis, 2004; Simonyan & Horwitz, 2011). However, while the emergence of this connection is likely to account for increased volitional control over the laryngeal muscles, it does not seem sufficient to account for the novel engagement of the respiratory musculature seen with LMC stimulation in humans. In non-human primates, specific vocal calls can be elicited by stimulation of either the periaqueductal grey (PAG) or the supra-genual anterior cingulate cortex (ACC: Jürgens & Pratt, 1979a, 1979b), while stimulation of the LMC in these species (i.e., area 6v) produces contraction of the laryngeal muscles *without* the respiratory drive for vocalization (Hast et al., 1974; Jürgens, 1974), in keeping with a non-vocal functionality of the LMC in these species. In contrast, the ACC projects to the PAG, which has descending projections to both the nucleus ambiguus and nucleus retroambiguus, which are a laryngeal and a respiratory brainstem nucleus, respectively (Jürgens & Müller-Preuss, 1977; Müller-Preuss & Jürgens, 1976; Vanderhorst, Terasawa, Ralston, & Holstege, 2000). The nucleus retroambiguus in turn projects both to laryngeal lower motor neurons in the nucleus ambiguus and to respiratory lower motor neurons in the spinal cord (VanderHorst, Terasawa, & Ralston, 2001; Vanderhorst, Terasawa, Ralston, & Holstege, 2000b), making it a likely target for the integration of these two components of vocalization (Holstege & Subramanian, 2016). The efferent pathways for vocalization are summarized graphically in Figure 3, which is supplemented with references in Supplementary Figure S1.

Unlike the situation in monkeys, stimulation of the human LMC regions does elicit vocalization, including the requisite expiratory drive (Breshears et al., 2015; Foerster, 1931; Penfield & Boldrey, 1937). Likewise, voluntary expiration (but not voluntary inspiration) leads to activation in the dorsal LMC in neuroimaging experiments (Loucks et al., 2007; Ramsey et al., 1993). Anatomical and neural changes to the system for voluntary control of respiration have been proposed as essential prerequisites for the evolution of speech (MacLarnon & Hewitt, 1999, 2004: Vaneechoutte, Munro & Verhaegen, 2011). In songbirds, nucleus RA, which is analogous to either the ventral or dorsal LMC in humans (Jarvis, 2004; Pfenning et al., 2014), projects to brainstem motor nuclei for both the respiratory and syringeal musculature in order to regulate vocalization (Wild, 1993; Wild, Kubke, & Mooney, 2009). From all of these observations, we hypothesize that humans may have evolved an as-yet-undiscovered efferent pathway from the (dorsal) LMC to the nucleus retroambiguus.

Insert Figure 3 about here

2.4 The "single vocal system" model

Before concluding this section about the structure of the larynx motor cortex, we would like to argue that the available data suggest that there is a single vocal system in the human brain that mediates all the vocal functions of human communication and expression, including speaking, singing, and the expression of emotions.

Myers (1976) observed that neurological trauma in human patients could selectively impact either speech or emotional vocalizations, while sparing the other. Several researchers later hypothesized that the human vocal system may consist of two functionally-distinct divisions: 1) the LMC for learned vocalizations, such as speech and song, and 2) the ACC/PAG axis for emotional vocalizations (Jürgens, 2009; Owren, Amoss, & Rendall, 2011; Simonyan & Horwitz, 2011). This hypothesis was based on the observation that emotional vocalizations in the monkey are regulated by a descending pathway originating in the ACC that projects to the PAG (reviewed in Jürgens, 2002, 2009), as well as on the correlation between comparative differences in the human and monkey LMCs and the differing vocal abilities of these species, suggesting a more specific involvement of the LMC in learned vocalizations, not least speech production. This two-pathway model predicts that i) spontaneous emotional vocalization activates the PAG, but neither cortical structure, ii) volitional emotional vocalization engages the ACC/PAG axis, but not the LMC, and iii) learned vocalization such as speech activates the LMC, but not the ACC/PAG axis.

However, due to developments in both neuroimaging and neurological research, the assignment of speech and emotional expression to separate vocal pathways has become less clear, leading to alternative proposals that all vocal functions are mediated by a single, common vocal system. Both the LMC and ACC are engaged during the production of both learned vocal patterns, such as speaking and singing (Brown et al., 2009), and emotional vocalizations (Aziz-Zadeh, Sheng, & Gheytanchi, 2010; Barrett, Pike, & Paus, 2004; Laukka, Åhs, Furmark, & Fredrikson, 2011; Wattendorf et al., 2013). We reported a direct comparison between volitional production of emotional vocalizations and the production of acoustically similar, but learned and non-emotional, vocalizations (Belyk & Brown, 2016). Using, fMRI, we found that the ACC/PAG axis and the LMC were each activated in a comparable manner by the production of both volitional emotional vocalizations and learned vocalizations. Together with previous studies (for spontaneous emotional vocalizations, see Barrett et al., 2004; Wattendorf et al., 2013), these findings suggest that a single integrated vocal-motor system, one that includes both the LMC and ACC/PAG descending pathways, may drive both learned vocalization and innate vocal expressions of emotion in humans.

An analysis of the neurological literature further questions the attribution of speech and emotional vocalization to separate LMC and ACC/PAG pathways, respectively. This two-pathway model would predict that lesions to the LMC would selectively affect propositional speech (as in motor aphasia), while lesions to the ACC would selectively affect vocal expressions of emotion (as in motor aprosodia). Contrary to these predictions, neurological reports demonstrate that lesions affecting the LMC are a frequent cause of motor aprosodia (Guranski & Podemski, 2015; Ross, 1981; Ross & Mesulam, 1979). Importantly, these reports extend to both spontaneous and volitional expressions of emotion (House, Rowe, & Standen, 1987; Trauner, Ballantyne, Friedland, & Chase, 1996). Jürgens, Kirzinger, & von Cramon (1982) reported a notable case study in which a stroke affecting the middle cerebral artery, including the supply to the LMC but not ACC/PAG, resulted in complete mutism, with the patient being unable to vocalize even in response to painful stimuli. Hence, although the neurological literature continues to observe that speech and emotional expression may be differentially impacted following brain damage, as described by Myers (1976), it does not support the specific attribution of emotional vocalizations to the ACC/PAG and learned vocalizations to the LMC.

Such findings have led to the development of hypotheses in which the components of the vocal-motor system function interactively. Ackermann et al. (2014) argued that the two vocal pathways, while distinct, must interact to coordinate the simultaneous production of propositional speech and emotional expression. Ludlow (2015) argued for a still more integrated view of the vocal-motor system that also incorporates the control of swallowing. We further propose to conceptualize the vocal-motor system as a single network for coordinating the laryngeal and respiratory muscles during vocalizations, and that any differences between speech, song, and emotional expression may be related to the nature of the inputs that drive this network, for example from language areas in the case of speech or the limbic system in the case of emotional expression.

Numerous neuroimaging comparisons of speech and song have been carried out. Early brain imaging studies suggested that speaking and singing may engage distinct networks or a common network lateralized to opposite hemispheres (Jeffries, Fritz, & Braun, 2003; Riecker, Ackermann, Wildgruber, Dogil, & Grodd, 2000). However, these trends were not replicated (Brown, Martinez, & Parsons, 2006; Ozdemir, Norton, & Schlaug, 2006; Perry et al., 1999), and meta-analysis revealed that speaking and singing engage highly overlapping networks, particularly with regard to the motor system (Brown et al., 2009; Ozdemir et al., 2006). Zatorre & Baum (2012) noted that differences between the neural systems for speech and song were more likely to occur at levels other than the processing of auditory input or vocal output, since these peripheral mechanisms are common to both systems. To the extent that the LMC controls the three dimensions of laryngeal movement (adduction vs. abduction, stretching vs. relaxing, and upward vs. downward), it seems to do so in a comparable manner for both the relatively discrete pitch-transitions that occur in song (e.g., intervals, scales) and the relatively continuous pitch-transitions that occur in speech. The shared processing of vocal-motor planning for speech and song is compatible with evolutionary models that argue that speech and song evolved from a common ancestral system that embodied their shared features (Brown, 2000; Darwin, 1871; Mithen, 2005).

Insert Figure 4 about here

Figure 4 presents a conceptual model of the vocal-motor system as a common output system for vocal communication. While “language”, “emotion”, and “music” are unquestionably distinct types of systems for communicating social meaning, the behaviors through which they are conveyed (i.e., speaking, emotional vocalizations, and singing) are mediated by a common neuromotor system controlling the vocal apparatus. The single vocal system includes not only the larynx motor cortex, but an extended audiovocal network, including the superior temporal gyrus, inferior frontal gyrus, supplementary motor area, anterior cingulate cortex, putamen, and lateral cerebellum (Brown et al., 2009; Guenther, Ghosh, & Tourville, 2006; Guenther & Vladusich, 2012). In addition, speech and song very often come together in the form of songs with words. Some forms are musical, with the use of scaled pitches and discrete intervals, while others are more speech-like, as is seen in "parlando" forms of chanting found in many world cultures (Lomax, 1968). Overall, the current state of the literature suggests that a single vocal system drives both learned vocal behaviors, such as speech and song, and innate vocal behaviors, such as emotional vocalizations.

3. Comparative neuroscience of the larynx motor cortex

As mentioned in the opening section, there are two critical questions about the evolution of vocalization in humans that need to be addressed. The first is how the larynx motor cortex of humans evolved from a presumably non-vocal LMC precursor in ancestral species. The second is how humans acquired the capacity for vocal learning from an ancestral species that lacked this capacity. We will address the first question here and the second question in the section "Comparative neuroscience of vocal production learning" below.

3.1 The primate LMC

Research on the LMC of non-human primates, primarily macaques and squirrel monkeys, has revealed that the non-human LMC differs from that in humans in both its cortical location and degree of involvement in vocalization. First, the monkey LMC is not located in the primary motor cortex (area 4) but instead in the ventral premotor cortex (area 6v), just posterior to the monkey homologue of Broca's area (Hast et al., 1974; Hast & Milojkvic, 1966; Jürgens, 1974). The fact that this premotor LMC location occurs in both of the major lineages of monkeys (Old World and New World) strongly suggests that it represents the ancestral state of primates. Second, these same studies demonstrated that, while electrical stimulation of the monkey

LMC causes the laryngeal muscles to contract (Hast et al., 1974; Jürgens, 1974), it does not elicit the respiratory changes necessary to drive vocalization (Walker & Green, 1938), as it does in humans (Breshears et al., 2015; Penfield & Boldrey, 1937). Even more importantly, experimental lesions to the monkey LMC have little effect on spontaneous vocal behavior (Kirzinger & Jürgens, 1982), although recording studies have observed the firing of LMC neurons in preparation for conditioned, but not spontaneous, vocalizations (Coudé et al., 2011; Hage & Nieder, 2013). Overall, while the monkey LMC clearly appears to be a larynx-controlling region, it does not play a critical role in vocalization.

What about great apes? Lesion studies of great apes have ceased in recent decades due to the endangered status of these species. However, the data from several early studies suggest that the LMC in great apes is intermediate between the monkey and human LMC in both cortical location and involvement in vocalization. By far the most extensive physiological study is that of Leyton & Sherrington (1917), which mapped cortical motor functions in three species of great apes (chimpanzee, gorilla, and orangutan). The authors observed that a variety of laryngeal movements – including vocal-fold adduction and abduction, vertical laryngeal movement through engagement of the extrinsic laryngeal muscles, and sound emission, primarily in the form of grunting – could be elicited by stimulation of the anterior edge of the ventral precentral gyrus. Hines (1940) claimed to be able to elicit vocalization in one of three chimpanzees through electrical stimulation of this same area, although the nature of this vocalization was not described. Potentially related to a rudimentary vocal function of the LMC in chimpanzees, Kuypers (1958a) reported the existence of sparse corticobulbar axons from the ventral precentral gyrus making direct synaptic contact onto neurons in the nucleus ambiguus. The intermediate vocal phenotype of non-human great apes suggests that selection for increased vocal-motor control had already begun at the time of the last common ancestor of the great ape lineage, although this ability has evidently been further elaborated over the course of human evolution.

3.2 Models of human LMC evolution

3.2.1 Duplication and migration

Brown et al. (2008) proposed that, because the dorsal LMC that they and others (Loucks et al., 2007; Rödel et al., 2004) characterized in humans occurs in a markedly different location from the monkey LMC, the human area must have undergone an evolutionary migration from the ancestral location in the ventral premotor cortex in monkeys to its human location adjacent to the somatotopic lip area in M1. However, in light of the later observations of Bouchard et al. (2013) and Pfenning et al. (2014) that there are in fact two larynx motor areas in each hemisphere of the human brain, it is necessary to revise this proposal.

The data of Leyton & Sherrington (1917) suggest that a first step in this evolution was a posterior relocation of the LMC from the ventral premotor cortex in monkeys to the ventral precentral gyrus in apes. We hypothesize that this posterior migration continued throughout hominid evolution, resulting in the ventral LMC of the Rolandic operculum as the human homologue. Three lines of evidence suggest that the ventral LMC, rather than the dorsal LMC, is the human homologue of the non-human primate LMC. First, the ventral LMC is more proximate to the LMC location in non-human primates. Indeed, the location of the ventral LMC is consistent with a continued posterior-ward relocation of the LMC from the ventral premotor cortex in monkeys to the ventral precentral gyrus in apes to the Rolandic operculum in humans (see Figure

5). Second, the limited evidence that is currently available from electrical stimulation studies in apes and humans suggests a greater similarity of the vocal responses elicited by stimulation of the ape LMC and the ventral LMC – rather than the dorsal LMC – in humans. Stimulation of the LMC in apes elicits grunt-like sounds (Leyton & Sherrington, 1917). In humans, electrical stimulation close to the ventral LMC also elicits grunting sounds (Foerster, 1931; Penfield & Boldrey, 1937), whereas stimulation near the dorsal LMC elicits vowel-like vocalizations reminiscent of speech (Breshears et al., 2015; Penfield & Boldrey, 1937). Third, the activation and morphology of the somatosensory cortex immediately posterior to the dorsal LMC are affected by singing experience (Kleber, Veit, Birbaumer, Gruzelier, & Lotze, 2010; Kleber et al., 2016), further suggesting that the dorsal LMC, as compared to the ventral LMC, may have a greater association with characteristically-human vocal forms, such as speech and song.

Insert Figure 5 about here

Figure 5 presents a graphic summary of two evolutionary hypotheses that attempt to account for the evolution of the dual representation of the larynx in the human motor cortex. Both hypotheses consider the ventral LMC of Bouchard et al. (2013) to be the human homologue of the ape LMC. Importantly, both models consider the dorsal LMC to be a human novelty, one that may be related to the evolutionary emergence of human-specific vocal capacities, such as voluntary control of vocalization and vocal learning (Brown et al., 2008).

The “duplication + migration model” (Figure 5A) posits that the dorsal LMC evolved by duplication of motor areas with pre-existing vocalization-related – laryngeal and/or respiratory – functions, followed by a long-distance migration to its current position. Duplication of the ventral LMC and/or trunk motor cortex would have required relatively few changes in white matter pathways to achieve the vocal, glottal-closure, and respiratory functionality of the dorsal LMC, since at least some of the necessary efferent pathways would have already been present in the precursor region. However, the relatively large distance between either of the proposed precursor-areas (ventral LMC and/or respiratory motor cortex) and the position of the dorsal LMC implies a considerable migration of neuronal cell bodies over the course of evolution, or more specifically a displacement of the patterns of gene expression that drive the development of LMC cells. One important piece of evidence in support of this model is that the dorsal and ventral LMC regions share patterns of gene expression relative to the surrounding precentral gyrus (Pfenning et al., 2014). This strongly suggests that these brain regions have a common origin.

As an alternative, the “local duplication model” (Figure 5B) proposes that the dorsal LMC evolved by duplication of an adjacent non-vocal part of the motor cortex. This would require considerable reorganization of connectivity patterns to acquire the vocal functionality of the dorsal LMC, but would not require an extensive evolutionary migration of neuronal cell bodies. Under this model, the dorsal and ventral LMCs are not homologous to one another, but evolved independently. Several variants of this hypothesis have been previously proposed. Feenders et al. (2008) observed that the vocal-motor nuclei of songbirds are adjacent to non-vocal motor areas, and hypothesized that the avian vocal system may have evolved as a specialization of a pre-existing non-vocal motor system. Chakraborty & Jarvis (2015) further observed that parrots, which are the most accomplished avian vocal learners, possess nested “shell” and “core” vocal systems. They hypothesized that the vocal-motor “shell” arose by duplication of the “core”, which in turn arose by duplication from non-vocal motor areas. They further postulated that similar processes may have occurred during human evolution. Fitch (2011) similarly hypothesized that the human LMC may have evolved from the adjacent hand motor cortex, although we note that other non-vocal motor areas, such as the lip or jaw areas, are equally plausible as potential precursor regions. A hand-based origin might be consistent with gestural models of language origin, which argue that speech arose from a pre-existing manual communicative system (Arbib, 2012; Hewes et al., 1973), whereas a lip- or jaw-based origin might be consistent with articulatory models that argue that speech arose from the union of phonation with pre-existing mandibular oscillatory movements, such as lip smacking in non-human primates (Ghazanfar, Morrill, & Kayser, 2013; Ghazanfar, Takahashi, Mathur, & Fitch, 2012; MacNeilage & Davis, 2005). Further research on the ontogeny and molecular genetic profiles of the ventral and dorsal LMCs will be required to test these hypotheses.

3.2.2 Descent of the larynx

Our proposal that the dorsal LMC may be a novel human brain area raises questions about the mechanisms by which new brain areas are able to arise. One well-known mechanism that may result in neural specializations within the central nervous system is a modification to peripheral effectors and/or life-history conditions. For example, there has been a strong regression of the primary visual cortex in the naked mole rat that lives in total darkness. Likewise, there has been an expansion of the primary somatosensory cortex in the duck-billed platypus that has experienced an extensive proliferation of mechanoreceptors on its bill (Krubitzer & Stolzenberg, 2014). With regard to human evolution, one of the most notable peripheral changes related to vocalization is the descent of the larynx in humans compared to non-human primates (Fitch, 2000a; Nishimura, 2003, 2006, 2008). This structural change is thought to have liberated the tongue to increase the complexity of phonemic repertoires in humans (Fitch, 2000b; Lieberman et al., 1969; although see Fitch, de Boer, Mathur, & Ghazanfar, 2016). Humans have undergone a two-part descent of the larynx, the first of which is shared with chimpanzees but not monkeys, namely descent of the cartilaginous skeleton of the larynx relative to the hyoid bone (Nishimura, 2005, 2006). Although purely correlative, this two-part change to the structure of the larynx across primate species is suggestive of the hypothesized two-stage posterior-ward relocation of the LMC, first from the premotor cortex in monkeys to the border of precentral gyrus in apes, and then to the dual primary motor representations in humans. Whether laryngeal descent was the driving force for the reorganization of the LMC in humans, or whether it was a completely independent adaptation, is something that needs to be explored in future comparative studies of animal species with descended larynges.

Aside from humans, there are a number of mammalian species that have a permanently descended larynx (Fitch, 2009; Fitch & Reby, 2001), and yet these species lack the human proficiency for vocalization. Hence, if the LMC has undergone substantial reorganization in these species, then these changes may be related to peripheral changes in the vocal-tract position of the larynx per se, rather than to the emergence of vocal complexity and vocal learning. One observation that argues against a causal relationship between descent of the larynx and LMC reorganization in humans is the fact that adult human males undergo an additional descent of the larynx at puberty that does not occur in adult females (Fitch & Giedd, 1999), and yet there is no evidence that LMC positioning differs between the sexes, or that it differs between men and boys. The issue could be investigated through brain imaging studies that compare the location of the LMCs between the sexes before and after puberty.

The somatotopic location of the dorsal LMC next to the lip representation – as well as its dual functionality for expiration and phonation – is surprising since its expected location would be adjacent to the neck and pharynx, as based on the mammalian body plan and the organization of the cranial nerve nuclei that innervate these muscles. The organization of the motor homunculus shows features of both continuity and discontinuity with regard to the body. On the one hand, effectors that are close together in the body tend to be proximate to one another in the motor cortex. To a general approximation, the superior-to-inferior structure of the face and oral tract is represented in a systematic manner along the dorsal-to-ventral extent of the orofacial part of the motor cortex, and to a rough approximation in the corresponding cranial nerve nuclei as well. However, if one progresses inferiorly along the body from the head to the thorax, one sees a significant discontinuity in homuncular organization, such that the trunk representation is located a great distance away from the head, in the most dorsal and medial part of the motor cortex, with the upper limb interceding between the head and thorax.

One hypothesis of LMC reorganization is that descent of the larynx toward the thorax led to a dorsal migration of the duplicate larynx representation in the direction of the trunk representation in M1. In addition, while the expiratory muscles of monkeys are located in the trunk region of the motor cortex, humans seem to have a double representation of these muscles, with one occurring in the trunk region (Colebatch et al., 1991) and the other occurring in the dorsal LMC region (Loucks et al., 2007; Ramsay et al., 1993; Simonyan, Saad, Loucks, Poletto, & Ludlow, 2007). While the trunk area that is shared between monkeys and humans is involved in both inspiration and expiration, the LMC area that is unique to humans seems to be more associated with expiration and vocalization than inspiration (although see the data on sniffing in Simonyan et al., 2007).

3.2.3 Brachiomotor confluence

Another evolutionary hypothesis about the unusual location of the dorsal LMC in the human brain is based on the idea that the nucleus ambiguus is one of three branchiomotor nuclei derived from the ancestral vertebrate system for innervating the gill arches of fish (Chandrasekhar, 2004; Guthrie, 2007). The other two are the trigeminal motor nucleus that controls the jaw muscles and the facial motor nucleus that controls the lip muscles (and other facial muscles). Hence, it is possible that the dorsal LMC's location achieved a "cortical confluence" of the three branchiomotor systems for the larynx, jaw and lips, respectively, as might be consistent with a mandibular-oscillation model of speech evolution (Ghazanfar et al., 2012, 2013; MacNeilage & Davis, 2005). This idea is supported by the fact that the trigeminal motor nucleus, facial motor nucleus, and nucleus ambiguus are organized as a single rostro-caudal cell column in the brain stem (Finger, 1993).

4. Comparative neuroscience of vocal production learning

Having discussed the anatomy and neurophysiology of the larynx motor cortex in humans as well as models of its evolution across the primate order, we would now like to discuss the notable behavioral phenotype culminating from this evolution, namely the capacity for vocal learning. While most animals have the ability to vocalize, they vary dramatically in their capacity for vocal learning, of which there are two major types. *Vocal usage learning* (Janik & Slater, 2000; Petkov & Jarvis, 2012) refers to the ability to learn when to produce vocalizations from an existing, generally innate, repertoire. For example, individuals may refrain from vocalizing or may vocalize deceptively, depending on the social composition of their audience (Fitch & Hauser, 2002; Loh et al., 2016; Munn, 1986; Townsend & Zuberbuhler, 2009). This ability is pervasive among primates (Koda, Oyakawa, Kato, & Masataka, 2007; Pierce, 1985). In contrast, *vocal production learning* refers to the ability to add new vocalizations to a repertoire, typically through vocal imitation. We will focus on vocal production learning in the remainder of this section, since this is an essential prerequisite for the evolution of speech and song in humans.

4.1 Vocal production learning in mammals

Vocal production learning is quite rare among animal species. While some great apes have been reported to produce novel sounds following extended exposure to humans, these sounds are usually produced in a non-vocal manner, through the use of sound sources such as lip smacking or whistling (Bergman, 2013; Hayes & Hayes, 1951; Hopkins, Taglialatela, & Leavens, 2007; Wich et al., 2009), rather than through the laryngeal sound source that underlies human vocalization. The captive chimpanzee Viki was able to learn articulatory and respiratory movements so as to produce a few English words, but never acquired the corresponding laryngeal movements (Hayes & Hayes, 1951). However, several case studies suggest that great apes may have a limited degree of control over the larynx. Wich et al. (2012) observed that certain call types were present in some groups of orangutans and absent in others, suggestive of a limited vocal-culture. Lameira et al. (2015) reported the production of a novel laryngeal sound in one captive orangutan. The captive gorilla Koko was observed to make one novel vocal sound, although this fell short of her more extensive repertoire of novel voiceless sounds (Perlman & Clark, 2015). Taken together, these findings demonstrate a limited capacity for vocal learning in great apes beyond the abilities of monkeys, but not approaching the abilities of human infants (Kuhl & Meltzoff, 1996).

Vocal production learning is more pronounced in three lineages of birds (discussed in the next section) and to some extent in several species of mammals, including, elephants (Poole, Tyack, Stoeger-Horwath, & Watwood, 2005; Stoeger et al., 2012), cetaceans (Janik, 2014; King & Sayigh, 2013; Noad, Cato, Bryden, Jenner, & Jenner, 2000), some species of bat (Knörnschild, Nagy, Metz, Mayer, & von Helversen, 2010; Vernes, 2016) and pinniped (Ralls, Fiorelli, & Gish, 1985; Reichmuth & Casey, 2014; Sanvito, Galimberti, & Miller, 2007; Schusterman & Feinstein, 1965).

Vocal production learning is not simply a motor capacity, but a *sensorimotor* mechanism that permits a perceived sound to be converted into a set of motor commands that can reproduce that sound, as in the matching of a heard pitch with the voice. Hence, one cannot create hypotheses about vocal imitation without giving serious consideration to the auditory mechanisms that allow an imitated object to be perceived to begin with. While direct corticobulbar connectivity from M1 to the nucleus ambiguus may be a necessary condition for vocal learning to evolve (Fitch, Huber & Bugnar, 2010), an understanding of the neural basis of vocal learning must also involve an elucidation of the *audiovocal* mechanisms that permit auditory percepts to be converted into the motor commands that vocally reproduce the perceived sound, for example through systems that mediate phonological working memory (Arboitiz, 2012), among other audiovocal capacities. Imitative learning is a sensorimotor, not just a motor, process.

An early model of vocal imitation in humans (Geschwind, 1970), largely derived from neurological observations, was predicated on the flow of auditory information from auditory areas in the posterior superior temporal gyrus (pSTG) to motor-planning areas in the IFG via the arcuate fasciculus (AF; see Fernandez-Miranda et al., 2015, and Glasser & Rilling, 2008, for structural analyses of the AF). Because the IFG does not contain upper motor neurons that project to the brainstem or spinal cord, information has to then be transmitted to vocal areas in M1 (such as the LMC) in order for vocal production to occur. This model has been taken to suggest that the audiomotor linkage established through the AF is both necessary and sufficient for vocal imitation to occur (critically discussed in Bernal & Ardila, 2009). Lesions restricted to the AF effectively disconnect auditory areas from the IFG, and can lead to a condition known as conduction aphasia, which is a paradoxical syndrome in which both speech comprehension and production are spared, but in which patients are unable repeat (i.e., imitate) heard utterances (Geschwind, 1970). In other words, patients have a specific impairment in the sensorimotor translation of auditory percepts into motor commands.

4.2 Songbirds as an animal model of vocal production learning

Although there has been very little research on the neural basis of vocal imitation or vocal learning in non-human mammals (reviewed in Arriaga & Jarvis, 2013), these abilities have been the subject of intense investigation in the three major lineages of vocal-learning birds, namely parrots, hummingbirds, and particularly songbirds (Nottebohm, 1972, Petkov & Jarvis, 2012). While songbirds are more phylogenetically distant from humans than are non-human primates, comparative analyses have demonstrated a marked degree of anatomical (Jarvis, 2004) and molecular genetic (Pfenning et al., 2014) similarity between the avian “song system” and the human audiovocal system, one that has led researchers to suggest that these species may have alighted on similar solutions to the problem of vocal learning through a process of convergent evolution (Jarvis, 2004). As with the human models based on neurological deficits in imitation, the birdsong literature has searched for neural mechanisms that permit target sounds to be mapped onto the motor commands that reproduce them, although with the experimental advantage of targeted neural lesions, compared to the idiosyncratic natural lesions that are the subject of neurological research in humans.

The avian song system consists of two interconnected pathways: a descending vocal-motor pathway and a forebrain-striatal loop (Jarvis et al., 2005). Both pathways receive input from the HVC, which may be related to a bird's repertoire of learned songs (Nottebohm, Stokes, & Leonard, 1976; Ward, Nordeen, & Nordeen, 1998). Unlike the descending motor pathway and forebrain striatal loop, the HVC may not correspond to any structure in the human brain (Pfenning et al., 2014). The descending pathway consists of the robust nucleus of the arcopallidum (RA) – which is the proposed analogue of the human LMC – although the vocal organ of the bird is not the larynx but instead the syrinx, which is innervated by RA via a direct projection to the hypoglossal nucleus, which itself gives rise to motor fibers that innervate the muscles of the syrinx. The forebrain-striatal loop consists of three structures: area X – the analogue of the human anterior striatum – the dorsolateral nucleus of the medial thalamus (DLM), and the lateral magnocellular nucleus of the anterior nidopallium (LMAN).

While lesions to the descending vocal-motor pathway profoundly disrupt song production (Nottebohm, Stokes, & Leonard, 1976), lesions to the forebrain-striatal loop disrupt vocal imitation and song learning, but spare the production of songs that have already been learned (Bottjer, Miesner, & Arnold, 1984; Brainard, 2004; Fee & Goldberg, 2011; Sohrabji, Nordeen, & Nordeen, 1990). Neurophysiological evidence suggests that neurons along the forebrain-striatal loop compute causal inverse models that map target sounds onto the motor commands that reproduce them (Giret, Kornfeld, Ganguli, & Hahnloser, 2014). These findings place processes critical to the sensorimotor aspect of vocal imitation within area X and related structures, leading to the hypothesis that analogous pathways may play a similar role in the human brain (Jarvis, 2007), especially the corticostriatal motor loop (Alexander & Crutcher, 1990).

We used fMRI to test this hypothesis by directly comparing imitative and non-imitative (i.e., pre-learned) vocal production of simple melodies in humans (Belyk et al., 2016). While both imitative and non-imitative vocalization engaged an identical set of motor and sensory brain areas, vocal imitation preferentially engaged the corticostriatal network, including the dorsal and ventral LMC, supplementary motor area, and – most notably – the striatum. This result presented a striking parallel to the network predicted by analogy with songbirds, supporting the similarity of function between the human striatum and songbird area X in forming inverse models of auditory targets. This demonstrates that, just as in songbirds, the human striatum plays an important role in vocal imitation, leading us to hypothesize that it may contain evolutionarily novel larynx-controlling circuitry not found in other primates for computing inverse causal models that map auditory targets onto the vocal-motor programs that reproduce them, akin to mechanisms found in avian vocal learners (Giret et al., 2014).

Furthermore, the convergent evolution of the songbird and human vocal-motor systems suggests candidates for molecular genetic mechanisms on which evolution may have acted to produce the human vocal-motor system. Certain genes within the Plexin, Neuropilin, Semaphorin, and Cadherin gene families, among others, are differentially expressed in the vocal nuclei of birds that are capable of vocal production learning, relative to those that are not (Matsunaga & Okanoya, 2008, 2009a, 2009b; Pfenning et al., 2014). The molecules produced by Plexin, Neuropilin, and Semaphorin genes are a family of cell-surface receptor proteins and their ligands, which together guide developing axons to their targets within the central nervous system (Dickson, 2002; Takahashi et al., 1999; Tamagnone et al., 1999). Cadherins produce cell-cell adhesion molecules that contribute to the aggregation and sorting of cells to form functionally differentiated gray matter regions and white matter tracts that connect them within a functional network (Redies, 1995; Redies, 2000). Although there has been little in-vivo research in humans to link variation in these genes to the development of the vocal-motor system, imaging genetics studies have begun to demonstrate the plausibility of such an approach (Belyk, Kraft, & Brown, 2014; Rujescu et al., 2007).

5. Evolutionary models of the human audiovocal system

Having discussed the comparative neuroscience of both the vocal system and vocal production learning, the important remaining issue is about the relationship between the two, in particular whether these mechanisms evolved sequentially or simultaneously. Ackermann et al. (2014) proposed a two-stage, sequential model of speech evolution. They hypothesized that the motor cortex first developed direct corticobulbar connections with the nucleus ambiguus, permitting volitional and flexible control over the laryngeal muscles, followed by an independent evolutionary event that elaborated the cortical and striatal circuitry for vocal imitation and vocal production learning. However, the evolutionary separation of volitional control of the larynx from vocal production learning implies the existence of an ancestral species of primate with the ability to volitionally produce flexible and novel vocal patterns but the paradoxical inability to learn these vocalizations from conspecifics. It remains for this perspective to identify the selective advantage of the former ability in the absence of the latter. One possibility is that volitional control of the larynx may have evolved to regulate non-vocal laryngeal functions, such as swallowing, although it is unclear that humans differ from other primates in this regard. A more likely selective advantage might be that volitional control of the larynx allowed this hypothetical ancestral primate to selectively exaggerate the apparent size of its body in order to communicate a more dominant social rank (Pisanski, Cartei, McGettigan, Raine, & Reby, 2016).

We would like to consider an alternative perspective in which the evolutions of vocal-motor control and vocal learning are linked, rather than being independent. Comparative neuroscience has revealed a progressive modification of brain morphology throughout the audiovocal system across primate orders. For example, consider the link between auditory association cortex and the frontal lobe. The temporal lobe projects to the inferior frontal gyrus by way of the two divisions of the arcuate fasciculus (AF) and the extreme capsule fiber complex, sometimes referred to as the dorsal and ventral language pathways, respectively (Friederici, 2012; Perani et al., 2011). Both of these pathways exist in at least a rudimentary form in monkeys, despite the poor vocal-motor abilities of these species (Mars et al., 2016; Petrides & Pandya, 2009; Thiebaut de Schotten, Dell'Acqua, Valabregue, & Catani, 2012). Rilling et al. (2008, 2012), using diffusion tensor imaging in living animals, demonstrated that the AF shows a progressive increase in size and target complexity from monkeys to chimpanzees to humans. A rudimentary AF exists in monkeys even though monkeys have a non-vocal LMC and poor vocal-learning abilities (Hast & Milojkvic, 1966; Jürgens, 1974). Temporo-frontal connectivity along the dorsal and ventral pathways is thus an ancestral neuroanatomical feature of primates that predates the capacity for vocal imitation and vocal production learning, although the expansion of the AF in great apes, and humans in particular, appears to correlate with the capacity for vocal production learning. Likewise, the posterior migration of the dorsal LMC in humans appears to have been accompanied by increased connectivity with structures in the parietal lobe (Kumar et al., 2016). This suggests that there has been a confluence of multiple neural changes to the vocal system over the course of primate evolution, rather than any single adaptation that has precipitated changes in human communicative abilities.

Figure 6 provides a summary of comparative research demonstrating local expansions in the AF and IFG, along with changes to the location, efferent projections, and neurophysiology of the LMC in primates (Supplementary Figure S2 provides the same figure with supporting references placed directly onto the figure). In each case, there is a neurophenotypic continuum from monkeys to great apes to humans that correlates with time since a last common ancestor and with increasing vocal and imitative abilities of these species. Looking beyond the primate order, Petkov & Jarvis (2012) proposed that a similar continuum of vocal production learning may usefully describe the abilities of other vertebrates. Their vocal-learning spectrum may correlate with the neurophenotypes that we have listed. For example, the ultrasonic vocalizations of mice are songlike (Holy & Guo, 2005), have a degree of acoustic flexibility suggestive of limited vocal production learning (Arriaga & Jarvis, 2013, although see contradictory evidence in Hammerschmidt et al., 2012, 2015, Kikusui et al., 2011), and are mediated by a primary-motor LMC with a sparse population of direct corticobulbar efferents (Arriaga et al., 2012), akin to those found in great apes (Kuypers, 1958a). Identifying the selective advantages of any one of these evolutionary changes in the absence of the others remains a fundamental challenge for a sequential view of vocal-motor evolution.

Insert Figure 6 about here

Figure 7 describes a novel alternative hypothesis of holistic vocal-motor-system evolution that attempts to tie together the previous discussion of the larynx motor cortex with the current discussion of vocal learning. The left panel presents the presumed ancestral state of primates, with a rudimentary AF connecting auditory areas with the IFG, and the latter projecting to a non-vocal LMC. The right panel presents a model of *brain pathway duplication* that links together the various morphological expansions throughout this system that were summarized in Figure 6, with the added proposal that these expanded pathways converged on a novel part of the cortex that resulted from the disproportionate expansion of the IFG during human evolution (Schenker et al., 2010). Under this hypothesis, this subregion of the IFG (most likely part of area 44) evolved as an “audiovocal hub” to integrate newly-evolved sensory, motor, and sensorimotor pathways in humans. In particular, we propose that this region integrated three critical facets of audiovocal connectivity during human brain evolution (Figure 7): 1) it acquired innervation from newly-evolved fibers during AF expansion, 2) it developed a novel, human-specific projection to the newly-evolved dorsal LMC to mediate sensorimotor control of vocalization (a pathway yet to be identified anatomically in the human brain), and 3) it interfaced with the basal ganglia to perform inverse-modeling operations critical to vocal imitation (not shown in the figure), as in songbirds. Just as we proposed in Figure 5 that the dorsal LMC may have evolved by duplication of other motor-cortical areas, here we propose that this process may have been part of a broader duplication of the human audiovocal system from the ancestral, non-vocal audiomotor system of non-human primates, as schematized by the circuit of purple-colored areas and pathways in Figure 7. According to this proposal, voluntary control of the vocal apparatus and vocal production learning *co-evolved* during human evolution, rather than evolving in a piecemeal, sequential manner.

Insert Figure 7 about here

6. Open questions about the origins of the vocal brain

Having reviewed both classic and recent findings about the vocal brain in humans, as well as having discussed evolutionary models of its origin, we would like to conclude by highlighting a number of remaining gaps in our knowledge about this system and its origins. Where possible, we outline hypotheses that can drive future research.

6.1 Regarding the anatomy and neurophysiology of the larynx motor cortex

1. In light of recent observations that the human motor homunculus contains two distinct laryngeal representations per hemisphere, we hypothesize that the dorsal LMC and ventral LMC may have distinct functional roles in controlling the larynx during both vocalization and airway protection.
2. If the dorsal and ventral LMC do have distinct functional roles, we hypothesize that they may have different connectomes, including both afferent and efferent connections.
3. Based on the novel property of the dorsal LMC in integrating laryngeal and respiratory motor control, we hypothesize that the corticobulbar efferents of the dorsal LMC may include direct or indirect projections to the nucleus retroambiguus that are absent in other primates.

4. We have outlined the absence of knowledge, both theoretical and empirical, regarding communication between the left and right LMCs across the corpus callosum in humans. Based on the synchronous and symmetrical nature of vocal-fold movements in normal laryngeal functioning, we hypothesize that the laryngeal motor corpus callosum operates on mechanisms other than the inter-hemispheric inhibition that characterizes the comparable pathways for the limbs in favor of alternative mechanisms that promote movement symmetry of the two vocal folds.
5. We have reiterated an hypothesis discussed elsewhere (Belyk & Brown, 2016; Ludlow, 2015) that a singular and integrated vocal-motor system, one that includes the dorsal LMC, ventral LMC and ACC, controls all functions of the larynx, including speech, song, emotional vocalization, and airway protection.

6.2 Regarding the evolution of the larynx motor cortex

6. We hypothesize that the ventral LMC is the human homologue of the non-human primate LMC. We also propose that the dorsal LMC may be a human novelty, one causally related to the evolution of speech and song (Brown et al., 2008).
7. We hypothesize that the dorsal LMC may have emerged by duplication of the ventral LMC and/or expiratory motor cortex, followed by migrations to its current location, perhaps driven by a "branchiomotor confluence" in the motor cortex.

8. We reiterated an alternative hypothesis that the dorsal LMC may have evolved independently of the ventral LMC, emerging by duplication of a non-laryngeal part of the motor cortex. Variants of this hypothesis have been proposed elsewhere (Feenders et al., 2008; Fitch, 2011).
9. We are agnostic with regard to whether the emergence of the LMC was driven by peripheral changes to the location of the larynx in the human vocal tract. However, we have outlined this hypothesis for completeness and have proposed conditions under which it could be tested.

6.3 Regarding the evolution of vocal production learning

10. In light of recent corroborating evidence, we reiterate the hypothesis of Jarvis (2004) that the avian song-system and the human audiovocal system may have alighted on similar mechanisms for vocal learning via convergent evolution, and point to the need to integrate this perspective with comparative evidence in the primate lineage. A likely focal point of such convergence is the striatum and its purported contribution to the formation of inverse causal models during vocal imitation.

6.4 Regarding the evolution of the vocal-motor system

11. We hypothesize that a subregion of the IFG may have emerged as an audiovocal hub during human evolution, serving as the recipient for novel auditory fibers from the human-expanded arcuate fasciculus, as an interface with the vocal-motor basal ganglia, and as a source of a novel projection to the newly-evolved dorsal LMC.

12. Finally, we propose that, if brain pathway duplication can be confirmed as the mechanism of vocal-system evolution in songbirds (as proposed by Chakraborty & Jarvis, 2015), then a similar mechanism may provide a parsimonious model for the simultaneous emergence of voluntary control of vocalization and the capacity for vocal learning in humans.

7. Acknowledgments

This work was funded by a grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada to S.B. The article benefitted greatly from a detailed reading of an earlier version of the manuscript by Erich Jarvis as well as his extensive suggestions for improvement. We thank Sonja Kotz for a critical reading of the manuscript.

References

Ackermann, H., Hage, S. R., & Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behavioral and Brain Sciences*, *37*, 529–604.

Ardran, G., & Kemp, F. (1952). The protection of the laryngeal airway during swallowing. *British Journal of Radiology, 25*, 406–416.

Alexander, G. E., & Crutcher, M. D. (1990). Functional architecture of basal ganglia circuits: Neural substrates of parallel processing. *Trends in Neurosciences*, *13*(7), 266–271.

Arbib, M. A. (2012). *How the brain got language: The mirror system hypothesis*. Oxford: Oxford University Press.

Arboitiz, F. (2012). Gestures, vocalizations, and memory in language origins. *Frontiers in Evolutionary Neuroscience*, *4*, 1-15.

Arriaga, G., Zhou, E. P., & Jarvis, E. D. (2012). Of mice, birds, and men: The mouse ultrasonic song system and vocal behavior. *PLoS ONE, 72*(7), 3859. doi:10.1371/journal.pone.0046610

Arriaga, G., & Jarvis, E. D. (2013). Mouse vocal communication system: Are ultrasounds learned or innate? *Brain and Language*, *124*(1), 96–116. doi:10.1016/j.bandl.2012.10.002

Assaf, Y., & Pasternak, O. (2008). Diffusion tensor imaging (DTI)-based white matter mapping in brain research: A review. *Journal of Molecular Neuroscience, 34*(1), 51–61. doi:10.1007/s12031-007-0029-0

Aziz-Zadeh, L., Sheng, T., & Gheytanchi, A. (2010). Common premotor regions for the perception and production of prosody and correlations with empathy and prosodic ability. *PLoS ONE*, *5*(1), e8759. doi:10.1371/journal.pone.0008759

Barbizet, J. (1958). Yawning. *Journal of Neurology, Neurosurgery, and Psychiatry*, *21*, 203–208.

Barrett, J., Pike, G. B., & Paus, T. (2004). The role of the anterior cingulate cortex in pitch variation during sad affect. *European Journal of Neuroscience*, *19*(2), 458–464.

Behrens, T. E. J., Berg, H. J., Jbabdi, S., Rushworth, M. F. S., & Woolrich, M. W. (2007). Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *Neuroimage, 34*(1), 144–155. doi:10.1016/j.neuroimage.2006.09.018

Belyk, M., & Brown, S. (2014). Somatotopy of the extrinsic laryngeal muscles in the human sensorimotor cortex. *Behavioural Brain Research*, *270*, 364–371. doi:10.1016/j.bbr.2014.05.048

Belyk, M., & Brown, S. (2016). Pitch underlies activation of the vocal system during affective vocalization. *Social Cognitive and Affective Neuroscience*, 11, 1078–1088. doi:10.1093/scan/nsv074

Belyk, M., Kraft, S. J., & Brown, S. (2014). PlexinA polymorphisms mediate the developmental trajectory of human corpus callosum microstructure. *Journal of Human Genetics*, *60*, 147–150.

Belyk, M., Pfordresher, P. Q., Liotti, M., & Brown, S. (2016). The neural basis of vocal pitch imitation in humans. *Journal of Cognitive Neuroscience*, *28*(4), 621-635. doi:10.1162/jocn

Bergman, T. J. (2013). Speech-like vocalized lip-smacking in geladas. *Current Biology*, *23*(7), R268–R2689. doi:10.1016/j.cub.2013.02.038

Bernal, B., & Ardila, A. (2009). The role of the arcuate fasciculus in conduction aphasia. *Brain*, *132*, 2309–2316.

Bottjer, S., Miesner, E., & Arnold, A. (1984). Forebrain lesions disrupt development but not maintenance of song in passerine birds. *Science*, *224*(4651), 901–903.

Bouchard, K. E., Mesgarani, N., Johnson, K., & Chang, E. F. (2013). Functional organization of human sensorimotor cortex for speech articulation. *Nature*, *495*(21), 327–332. doi:10.1038/nature11911

Brainard, M. S. (2004). Contributions of the anterior forebrain pathway to vocal plasticity. *Annals of the New York Academy of Sciences, 1016*, 377–394. doi:10.1196/annals.1298.042

Breshears, J. D., Molinaro, A. M., & Chang, E. F. (2015). A probabilistic map of the human ventral sensorimotor cortex using electrical stimulation. *Journal of Neurosurgery*. doi:10.3171/2014.11.JNS14889

Brodmann, K. (1909). *Localisation in the cerebral cortex*. (L. J. Garey, Ed.) (3rd ed.). New York: Springer.

Brown, S. (2000). The "musilanguage" model of music evolution. In N. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 271–300). Cambridge, MA: MIT Press.

Brown, S., Laird, A. R., Pfordresher, P. Q., Thelen, S. M., Turkeltaub, P., & Liotti, M. (2009). The somatotopy of speech: Phonation and articulation in the human motor cortex. *Brain and Cognition*, *70*(1), 31–41. doi:10.1016/j.bandc.2008.12.006

Brown, S., Martinez, M. J., & Parsons, L. M. (2006). Music and language side by side in the brain: A PET study of the generation of melodies and sentences. *European*

Journal of Neuroscience, *23*(10), 2791–2803.

Brown, S., Ngan, E., & Liotti, M. (2008). A larynx area in the human motor cortex. *Cerebral Cortex*, *18*(4), 837–845. doi:doi: 10.1093/cercor/bhm131

Buchthal, F. (1959). Electromyography of intrinsic laryngeal muscles. *Quarterly Journal of Experimental Physiology and Cognate Medical Sciences*, *44*(2), 137–148.

Chakraborty, M., & Jarvis, E. D. (2015). Brain evolution by brain pathway duplication. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1683), 20150056. doi:10.1098/rstb.2015.0056

Chandrasekhar, A. (2004). Turning heads: Development of vertebrate branchiomotor neurons. *Developmental Dynamics*, *229*, 143–161.

Chang, E. F., Niziolek, C. A., Knight, R. T., Nagarajan, S. S., & Houde, J. F. (2013). Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proceedings of the National Academy of Sciences, 110*(7), 2653–2658. doi:10.1073/pnas.1216827110

Christiner, M., & Reiterer, S. M. (2013). Song and speech: Examining the link between singing talent and speech imitation ability. *Frontiers in Psychology, 4*, 1–11. doi:10.3389/fpsyg.2013.00874

Cogan, G. B., Thesen, T., Carlson, C., Doyle, W., Devinsky, O., & Pesaran, B. (2014). Sensory-motor transformations for speech occur bilaterally. *Nature*, *507*(7490), 94–98. doi:10.1038/nature12935

Colebatch, J. G., Adams, L., Murphy, K., Martin, A. J., Lammertsma, A. A., Tochon-Danguy, H. J., … Guz, A. (1991). Regional cerebral blood flow during volitional breathing in man. *Journal of Physiology*, *443*(1), 91–103. doi:10.1007/s13398-

014-0173-7.2

Conant, D., Bouchard, K., & Chang, E. (2014). Speech map in the human ventral sensory-motor cortex. *Current Opinion in Neurobiology*, *24*, 63–67.

Coudé, G., Ferrari, P. F., Rodà, F., Maranesi, M., Borelli, E., Veroni, V., … Fogassi, L. (2011). Neurons controlling voluntary vocalization in the macaque ventral premotor cortex. *PLoS ONE*, *6*(2783), 1–10. doi:10.1371/journal.pone.0026822

Darwin, C. (1871). *The descent of man, and selection in relation to sex*. London: Murray.

Dickson, B. J. (2002). Molecular mechanisms of axon guidance. *Science*, *298*(5600), 1959–1964.

Fant, G. 1960. *The acoustic theory of speech production*. The Hague: Moulton.

Fernández-Miranda, J. C., Wang, Y., Pathak, S., Stefaneau, L., Verstynen, T., & Yeh, F. C. (2015). Asymmetry, connectivity, and segmentation of the arcuate fascicle in the human brain. *Brain Structure and Function*, *220*, 1665–1680.

Fee, M. S., & Goldberg, J. H. (2011). A hypothesis for basal ganglia-dependent reinforcement learning in the songbird. *Neuroscience, 198*, 152–170. doi:10.1016/j.neuroscience.2011.09.069

Finger, T. E. (1993). What's so special about special visceral? *Acta Anatomica*, *148*, 132-138.

Fischer, J., & Hammerschmidt, K. (2011). Ultrasonic vocalizations in mouse models for speech and socio-cognitive disorders: Insights into the evolution of vocal communication. *Genes, Brain and Behavior, 10*(1), 17–27. doi:10.1111/j.1601-183X.2010.00610.x

Fitch, W. T. (2000a). The evolution of speech: A comparative review. *Trends in*

Cognitive Sciences, *4*, 258–267.

Fitch, W. T. (2000b). The phonetic potential of nonhuman vocal tracts: Comparative cineradiographic observations of vocalizing animals. *Phonetica*, *57*, 205–218.

Fitch, W. T. (2009). Fossil cues to the evolution of speech. In R. Botha and C. Knight (Eds.) *The cradle of language* (pp. 112–134). Oxford: Oxford University Press.

Fitch, W. T., de Boer, B., Mathur, N., & Ghazanfar, A. A. (2016). Monkey vocal tracts are speech-ready. *Science Advances*, *2*, e1600723.

Fitch, W. T., & Giedd, J (1999). Morphology and development of the human vocal tract: A study using MRI. *Journal of the Acoustical Society of America*, *106*, 1511-1522.

Fitch, W. T., & Hauser, M. D. (2002). Unpacking "honesty": Vertebrate vocal production and the evolution of acoustic signals. In A. M. Simmons, A. N. Popper & R. R. Fay (Eds.), *Acoustic communication* (Vol. 16, pp. 65-137). New York: Springer.

Fitch, W. T., Huber, L., & Bugnyar, T. (2010). Social cognition and the evolution of language: Constructing cognitive phylogenies. *Neuron*, *65*, 795-814.

Fitch, W. T. (2011). The evolution of syntax: An exaptationist perspective. *Frontiers in Evolutionary Neuroscience*, *3*(DEC), 1–12. doi:10.3389/fnevo.2011.00009

Fitch, W. T., & Reby, D. (2001). The descended larynx is not uniquely human. *Proceedings of the Royal Society*, *B*, 268, 1669–1675.

Fling, B. W., Benson, B. L., & Seidler, R. D. (2013). Transcallosal sensorimotor fiber tract structure-function relationships. *Human Brain Mapping*, *34*(2), 384–395. doi:10.1002/hbm.21437

Foerster, O. (1931). The cerebral cortex in man. *Lancet*, *218*(5632), 309–312.

Friederici, A. D. (2012). Language development and the ontogeny of the dorsal pathway. *Frontiers in Evolutionary Neuroscience*, *4*, 1–7.

Gay, E. C., Hirose, H., Strome, M., & Sawashima, M. (1972). Electromyography of the intrinsic laryngeal muscles during phonation. *Annals of Otology, Rhinology and Laryngology*, *81*(3), 401–410.

Geissmann, T. (2000). Gibbon songs and human music from an evolutionary perspective. In N. Wallin, B. Merker, and S. Brown (Eds.), *The origins of music* (pp. 103-123). Cambridge, MA: MIT Press.

Geschwind, N. (1970). The organization of language and the brain. *Science*, *170*(3961), 940–944.

Ghazanfar, A. A., Morrill, R. J., & Kayser, C. (2013). Monkeys are perceptually tuned to facial expressions that exhibit a theta-like speech rhythm. *Proceedings of the National Academy of Sciences*, *110*(5), 1959–1963. doi:10.1073/pnas.1214956110

Ghazanfar, A. A., Takahashi, D. Y., Mathur, N., & Fitch, W. T. (2012). Cineradiography of monkey lip-smacking reveals putative precursors of speech dynamics. *Current Biology*, *22*(13), 1176–1182.

Giret, N., Kornfeld, J., Ganguli, S., & Hahnloser, R. H. R. (2014). Evidence for a causal inverse model in an avian cortico-basal ganglia circuit. *Proceedings of the National Academy of Sciences*, *111*(16), 6063–6068. doi:10.1073/pnas.1317087111

Glasser, M. F., & Rilling, J. K. (2008). DTI tractography of the human brain's language pathways. *Cerebral Cortex*, *18*(11), 2471–2482. doi:10.1093/cercor/bhn011

Grabski, K., Lamalle, L., Vilain, C., Schwartz, J.-L., Vallée, N., Tropres, I., … Sato,

M. (2012). Functional MRI assessment of orofacial articulators: Neural correlates of lip, jaw, larynx, and tongue movements. *Human Brain Mapping*, *33*(10), 2306–2321. doi:10.1002/hbm.21363

Gracco, V. L., & Löfqvist, A. (1994). Speech motor coordination and control: Evidence from lip, jaw, and laryngeal movements. *Journal of Neuroscience, 14*, 6585–6597.

Gray, H. (1918). *Anatomy of the human body.* (20th ed.). Philidelphia: Lea & Febiger.

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, *96*(3), 280–301. doi:10.1016/j.micinf.2011.07.011.Innate

Guenther, F. H., & Vladusich, T. (2012). A neural theory of speech acquisition and production. *Journal of Neurolinguistics*, *25*(5), 408–422. doi:10.1016/j.jneuroling.2009.08.006

Guthrie, S. (2007). Patterning and axon guidance of cranial motor neurons. *Nature Reviews Neuroscience*, *8*, 859–871.

Guranski, K., & Podemski, R. (2015). Emotional prosody expression in acoustic analysis in patients with right hemisphere ischemic stroke. *Neurologia i Neurochirurgia Polska, 49*(2), 113–120. doi:10.1016/j.pjnns.2015.03.004

Hage, S. R., & Nieder, A. (2013). Single neurons in monkey prefrontal cortex encode volitional initiation of vocalizations. *Nature Communications, 4*, 2409. doi:10.1038/ncomms3409

Hammerschmidt, K., Reisinger, E., Westekemper, K., Ehrenreich, L., Strenzke, N., & Fischer, J. (2012). Mice do not require auditory input for the normal development of their ultrasonic vocalizations. *BMC Neuroscience, 13*(1), 40. doi:10.1186/1471-2202-13-40

Hammerschmidt, K., Schreiweis, C., Minge, C., Pääbo, S., Fischer, J., & Enard, W. (2015). A humanized version of Foxp2 does not affect ultrasonic vocalization in adult mice. *Genes, Brain, and Behavior, 14*, 583–590. doi:10.1111/gbb.12237

Hast, M. H., Fischer, J. M., Wetzel, A. B., & Thompson, V. E. (1974). Cortical motor representation of the laryngeal muscles in *Macaca mulatta*. *Brain*, *73*, 229–240.

Hast, M. H., & Milojkvic, R. (1966). The response of the vocal folds to electrical stimulation of the inferior frontal cortex of the squirrel monkey. *Acta Oto-Laryngologica*, *61*, 196–204. doi:10.3109/00016486609127056

Hayes, K. J., & Hayes, C. (1951). The intellectual development of a home-raised chimpanzee. *Proceedings of The American Philosophical Society*, *95*, 105–109.

Hewes, G. W., Andrew, R. J., Carini, L., Choe, H., Gardner, R. A., Kortlandt, A., … Wescott, R. W. (1973). Primate communication and the gestural origin of language. *Current Anthropology*, *14*, 5–24. doi:10.1086/201401

Hines, M. (1940). Movements elicited from precentral gyrus of adult chimpanzees by stimulation with sine wave currents. *Journal of Neurophysiology*, *3*, 442–466.

Hofer, S., & Frahm, J. (2006). Topography of the human corpus callosum revisited: Comprehensive fiber tractography using diffusion tensor magnetic resonance imaging. *Neuroimage*, *32*(3), 989–994. doi:10.1016/j.neuroimage.2006.05.044

Hollien, H., & Moore, P. (1960). Measurements of the vocal folds during changes in pitch. *Journal of Speech, Language and Hearing Research*, *3*, 157–165.

Holstege, G., & Subramanian, H. H. (2016). Two different motor systems are needed to generate human speech. *Journal of Comparative Neurology*, *524*(8), 1558–

1577. doi:10.1002/cne.23898

Holy, T. E., & Guo, Z. (2005). Ultrasonic songs of male mice. *PLoS Biology, 3*(12), 1–10. doi:10.1371/journal.pbio.0030386

Hopkins, W. D., Taglialatela, J., & Leavens, D. a. (2007). Chimpanzees differentially produce novel vocalizations to capture the attention of a human. *Animal Behaviour*, *73*(2), 281–286. doi:10.1016/j.anbehav.2006.08.004

House, A., Rowe, D., & Standen, P. J. (1987). Affective prosody in the reading voice of stroke patients. *Journal of Neurology, Neurosurgery, and Psychiatry, 50*(7), 910–912. doi:10.1136/jnnp.50.7.910

Isshiki, N., Tanabe, M., Ishizaka, K., & Broad, D. (1977). Clinical significance of asymmetrical vocal cord tension. *Annals of Otology, Rhinology and Laryngology*, *86*, 58–66.

Iwatsubo, T., Kuzuhara, S., & Kanemitsu, A. (1990). Corticofugal projections to the motor nuclei of the brainstem and spinal cord in humans. *Neurology*, *40*, 309–312.

Janik, V. M. (2014). Cetacean vocal learning and communication. *Current Opinion in Neurobiology, 28*, 60–65. doi:10.1016/j.conb.2014.06.010

Janik, V. M., & Slater, P. J. B. (2000). The different roles of social learning in vocal communication. *Animal Behaviour*, *60*(1), 1–11. doi:10.1006/anbe.2000.1410

Jarvis, E. D. (2004). Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences, 1016*, 749–777.

Jarvis, E. (2007). Neural systems for vocal learning in birds and humans: A synopsis. *Journal of Ornithology*, *148*(1), 35–44.

Jarvis, E., Güntürkün, O., Bruce, L., Csillag, A., Karten, H., Kuenzel, W., … Butler,

A. (2005). Avian brains and a new understanding of vertebrate brain evolution. *Nature Reviews Neuroscience*, *6*, 151–159.

Jeffries, K. J., Fritz, J. B., & Braun, A. R. (2003). Words in melody: An $H_2{}^{15}O$ PET study of brain activation during singing and speaking. *NeuroReport, 14*(5), 749–754. doi:10.1097/01.wnr.0000066198.94941.a4

Judaš, M., & Cepanec, M. (2010). Oskar Vogt: The first myeloarchitectonic map of the human frontal cortex. *Translational Neuroscience*, *1*(1), 72–94. doi:10.2478/v10134-010-0005-z

Jürgens, U. (1974). On the elicitability of vocalization from the cortical larynx area. *Brain Research*, *81*(3), 564–566.

Jürgens, U. (1976). Projections from the cortical larynx area in the squirrel monkey. *Experimental Brain Research, 25*(4), 401–411. doi:10.1007/BF00241730

Jürgens, U. (2002). Neural pathways underlying vocal control. *Neuroscience & Biobehavioral Reviews*, *26*(2), 235–258.

Jürgens, U. (2009). The neural control of vocalization in mammals: A review. *Journal of Voice*, *23*(1), 1–10. doi:10.1016/j.jvoice.2007.07.005

Jürgens, U., & Ehrenreich, L. (2007). The descending motorcortical pathway to the laryngeal motoneurons in the squirrel monkey. *Brain Research*, *1148*, 90–95. doi:10.1016/j.brainres.2007.02.020

Jürgens, U., Kirzinger, A., & von Cramon, D. (1982). The effects of deep-reaching lesions in the cortical face area on phonation: A combined case report and experimental monkey study. *Cortex, 18*, 125–139.

Jürgens, U., & Müller-Preuss, P. (1977). Convergent projections of different limbic vocalization areas in the squirrel monkey. *Experimental Brain Research, 29*(1), 75–83. doi:10.1007/BF00236876

Jürgens, U., & Pratt, R. (1979a). Role of the periaqueductal grey in vocal expression of emotion. *Brain Research*, *167*, 367–378.

Jürgens, U., & Pratt, R. (1979b). The cingular vocalization pathway in the squirrel monkey. *Experimental Brain Research*, *510*, 499–510.

Keller, S. S., Roberts, N., & Hopkins, W. (2009). A comparative magnetic resonance imaging study of the anatomy, variability, and asymmetry of Broca's area in the human and chimpanzee brain. *Journal of Neuroscience, 29*(46), 14607–14616. doi:10.1523/JNEUROSCI.2892-09.2009

Kempster, G. B., Larson, C. R., & Kistler, M. K. (1988). Effects of electrical stimulation of cricothyroid and thyroarytenoid muscles on voice fundamental frequency. *Journal of Voice*, *2*(3), 221–229. doi:10.1016/S0892-1997(88)80080-8

Kikusui, T., Nakanishi, K., Nakagawa, R., Nagasawa, M., Mogi, K., & Okanoya, K. (2011). Cross fostering experiments suggest that mice songs are innate. *PLoS ONE, 6*(3), 1–10. doi:10.1371/journal.pone.0017721

King, S., & Sayigh, L. (2013). Vocal copying of individually distinctive signature whistles in bottlenose dolphins. *Proceedings of the Royal Society B*, *280*, 1–9.

Kirzinger, A., & Jürgens, U. (1982). Cortical lesion effects and vocalization in the squirrel monkey. *Brain Research*, *233*, 299–315.

Kleber, B., Veit, R., Birbaumer, N., Gruzelier, J., & Lotze, M. (2010). The brain of opera singers: Experience-dependent changes in functional activation. *Cerebral Cortex, 20*(5), 1144–1152.

Kleber, B., Veit, R., Moll, C. V., Gaser, C., Birbaumer, N., & Lotze, M. (2016). Voxel-based morphometry in opera singers: Increased gray-matter volume in right somatosensory and auditory cortices. *Neuroimage, 133*, 477-483.

Knörnschild, M., Nagy, M., Metz, M., Mayer, F., & von Helversen, O. (2010). Complex vocal imitation during ontogeny in a bat. *Biology Letters*, *6*(2), 156–159. doi:10.1098/rsbl.2009.0685

Koda, H., Oyakawa, C., Kato, A., & Masataka, N. (2007). Experimental evidence for the volitional control of vocal production in an immature gibbon. *Behaviour*, *144*(6), 681–692. doi:10.1163/156853907781347817

Kochis-Jennings, K. A., Finnegan, E. M., Hoffman, H. T., Jaiswal, S., & Hull, D. (2014). Cricothyroid muscle and thyroarytenoid muscle dominance in vocal register control: Preliminary results. *Journal of Voice*, *28*(5), 652.e21-652.e29. doi: 10.1016/j.jvoice.2014.01.017

Krubitzer, L., & Stolzenberg, D. S. (2014). The evolutionary masquerade: Genetic and epigenetic contributions to the neocortex. *Current Opinion in Neurobiology*, *24*, 157–165.

Kryshtopava, M., Van Lierde, K., Meerschman, I., D'Haeseleer, E., De Moor, M., Vandemaele, P., … Claeys, S. (2017). Functional magnetic resonance imaging study of brain activity associated with pitch adaptation during phonation in healthy women without voice disorders. *Journal of Voice, 31,* 118.e21.

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation. *Journal of the Acoustical Society of America*, *100*(4), 2425–

2438.

Kumar, V., Croxson, P. L., & Simonyan, K. (2016). Structural organization of the laryngeal motor cortical network and its implication for evolution of speech production. *Journal of Neuroscience, 36*(15), 4170–4181.

Kuypers, H. G. J. M. (1958a). Some projections from the peri-central cortex to the pons and lower brain stem in monkey and chimpanzee. *Journal of Comparative Neurology*, *110*, 221–255.

Kuypers, H. G. J. M. (1958b). Corticobulbar connexions to the pons and lower brain-stem in man. *Brain*, *81*(3), 364–388. doi:10.1093/brain/81.3.364

Lameira, A. R., Hardus, M. E., Bartlett, A. M., Shumaker, R. W., Wich, S. A., & Menken, S. B. J. (2015). Speech-like rhythm in a voiced and voiceless orangutan call. *PLoS ONE*, *10*(1), e116136. doi:10.1371/journal.pone.0116136

Lang, I. M., Dana, N., Medda, B. K., & Shaker, R. (2002). Mechanisms of airway protection during retching, vomiting, and swallowing. *American Journal of Physiology - Gastrointestinal and Liver Physiology, 283*(3), G529–536. doi:10.1152/ajpgi.00062.2002

Laukka, P., Åhs, F., Furmark, T., & Fredrikson, M. (2011). Neurofunctional correlates of expressed vocal affect in social phobia. *Cognitive, Affective, & Behavioral Neuroscience*, *11*(3), 413–425. doi:10.3758/s13415-011-0032-3

Lieberman, P. H., Klatt, D. H., & Wilson, W. H. (1969) Vocal tract limitations on the vowel repertoires of rhesus monkeys and other nonhuman primates. *Science*, *164*, 1185–1187.

Leyton, S., & Sherrington, C. (1917). Observations on the excitable cortex of the chimpanzee, organ-utan, and gorilla. *Experimental Physiology*, *11*(2), 135–222.

Loh, K. K., Michael, P., William, H. D., Emmanuel, P., & Céline, A. (2016). Cognitive control of vocalizations in the primate ventrolateral-dorsomedial frontal (VLF-DMF) brain network. *Neuroscience & Biobehavioral Reviews*, 1–13. doi:10.1016/j.neubiorev.2016.12.001

Lomax, A. (1968). *Folk song style and culture*. New Brunswick, N.J. Transaction Books.

Loucks, T. M. J., Poletto, C. J., Simonyan, K., Reynolds, C. L., & Ludlow, C. L. (2007). Human brain activation during phonation and exhalation: Common volitional control for two upper airway functions. *Neuroimage*, *36*(1), 131–143. doi:10.1016/j.neuroimage.2007.01.049

Lowell, S. Y., & Story, B. H. (2006). Simulated effects of cricothyroid and thyroarytenoid muscle activation on adult-male vocal fold vibration. *Journal of the Acoustical Society of America*, *120*(1), 386–397. doi:10.1121/1.2204442

Ludlow, C. L. (2005). Central nervous system control of the laryngeal muscles in humans. *Respiratory Physiology & Neurobiology*, *147*(2-3), 205–222. doi:10.1016/j.resp.2005.04.015

Ludlow, C. L. (2015). Central nervous system control of voice and swallowing. *Journal of Clinical Neurophysiology*, *32*(4), 294–303.

MacLarnon, A., & Hewitt, G. (1999). The evolution of human speech. *American Journal of Physical Anthropology*, *109*, 341–363.

MacLarnon, A., & Hewitt, G. (2004). Increased breathing control: Another factor in the evolution of human language. *Evolutionary Anthropology*, *13*, 181–197.

MacNeilage, P. F., & Davis, B. L. (2005). The frame/content theory of evolution of speech: A comparison with a gestural-origins alternative. *Interaction Studies*, *6*(2), 73–99.

Mars, R. B., Foxley, S., Verhagen, L., Jbabdi, S., Sallet, J., Noonan, M. A. P., … Rushworth, M. F. S. (2016). The extreme capsule fiber complex in humans and macaque monkeys: A comparative diffusion MRI tractography study. *Brain Structure and Function*, *221*(8), 4059–4071. doi:10.1007/s00429-015-1146-0

Matsunaga, E., & Okanoya, K. (2008). Expression analysis of cadherins in the songbird brain: Relationship to vocal system development. *Journal of Comparative Neurology*, *508*(2), 329–342. doi:10.1002/cne.21676

Matsunaga, E., & Okanoya, K. (2009a). Evolution and diversity in avian vocal system: An Evo-Devo model from the morphological and behavioral perspectives. *Development Growth and Differentiation*, *51*(3), 355–367. doi:10.1111/j.1440-169X.2009.01091.x

Matsunaga, E., & Okanoya, K. (2009b). Vocal control area-related expression of neuropilin-1, plexin-A4, and the ligand semaphorin-3A has implications for the evolution of the avian vocal system. *Development, Growth & Differentiation*, *51*(1), 45–54. doi:10.1111/j.1440-169X.2008.01080.x

Merker, B., Morley, I., & Zuidema, W. (2015). Five fundamental constraints on theories of the origins of music. *Philosophical Transactions of the Royal Society B: Biological Sciences, 370*(1664), 20140095–20140095. doi:10.1098/rstb.2014.0095

Mithen, S. J. (2005). *The singing Neanderthals: The origins of music, language, mind and body*. London: Weidenfeld & Nicolson.

Miyaji, H., Hironaga, N., Umezaki, T., Hagiwara, K., Shigeto, H., Sawatsubashi, M., … Komune, S. (2014). Neuromagnetic detection of the laryngeal area: Sensoryevoked fields to air-puff stimulation. *Neuroimage, 88*, 162–169. doi:10.1016/j.neuroimage.2013.11.008

Müller-Preuss, P., & Jürgens, U. W. E. (1976). Projections from the “cingular” vocalization area in the squirrel monkey. *Brain Research, 103*, 29–43.

Munn, C. A. (1986). The deceptive use of alarm calls by sentinel species in mixed species flocks of neotropical birds. In R. W. Mitchell & N. S. Thompson (Eds.), *Deception: Perspectives on human and nonhuman deceit* (pp. 169-175). Albany, NY: State University of New York Press.

Myers, R. E. (1976). Comparative neurology of vocalization and speech: Proof of a dichotomy. *Annals of the New York Academy of Sciences*, *280*(1), 745–757.

Netz, J., Ziemann, U., & Hömberg, V. (1995). Hemispheric asymmetry of transcallosal inhibition in man. *Experimental Brain Research*, *104*(3), 527–533. doi:7589304

Nishimura, T. (2003). Comparative morphology of the hyo-laryngeal complex in anthropoids: Two steps in the evolution of the descent of the larynx. *Primates*, *44*, 41–49.

Nishimura, T. (2005). Developmental changes in the shape of the supralaryngeal vocal tract in chimpanzees. *American Journal of Physical Anthropology*, *126*, 193-204.

Nishimura, T. (2006). Descent of the larynx in chimpanzees: Mosaic and multiple-step evolution of the foundations for human speech. In T. Matauzawa, M. Tomonaga & M. Tanaka (Eds.) *Cognitive development in chimpanzees* (pp. 75-85). Tokyo: Springer.

Nishimura, T. (2008). Understanding the dynamics of primate vocalization and its implications for the evolution of human speech. In M. Masatak (Ed.) *The origins of language: Unraveling evolutionary forces* (pp. 111-131). Tokyo: Springer.

Noad, M. J., Cato, D. H., Bryden, M. M., Jenner, M. N., & Jenner, K. C. (2000). Cultural revolution in whale songs. *Nature*, *408*(6812), 537. doi:10.1038/35046199

Nottebohm, F. (1972). The origins of vocal learning. *American Naturalist*, *106*(947), 116–140.

Nottebohm, F., Stokes, T. M., & Leonard, C. M. (1976). Central control of song in the canary, Serinus canarius. *Journal of Comparative Neurology*, *165*(4), 457–486. doi:10.1002/cne.901650405

Olthoff, A., Baudewig, J., Kruse, E., & Dechent, P. (2008). Cortical sensorimotor control in vocalization: A functional magnetic resonance imaging study. *Laryngoscope*, *118*(11), 2091–2096. doi:10.1097/MLG.0b013e31817fd40f

Owren, M. J., Amoss, R. T., & Rendall, D. (2011). Two organizing principles of vocal production: Implications for nonhuman and human primates. *American Journal of Primatology*, *73*(6), 530–544. doi:10.1002/ajp.20913

Ozdemir, E., Norton, A., & Schlaug, G. (2006). Shared and distinct neural correlates of singing and speaking. *Neuroimage*, *33*(2), 628–635. doi:10.1016/j.neuroimage.2006.07.013

Pabst, F., & Sundberg, J. (1993). Tracking multi-channel electroglottograph measurement of larynx height in singers. *Logopedics Phonatrics Vocology*, *18*, 143–152. doi:10.3109/14015439309101360

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience, 6*(7), 674–681. doi:10.1038/nn1082

Patel, A. D. (2008). *Music, language and the brain*. Oxford: Oxford University Press.

Peck, K. K., Galgano, J. F., Branski, R. C., Bogomolny, D., Ho, M., Holodny, A. I., & Kraus, D. H. (2009). Event-related functional MRI investigation of vocal pitch variation. *Neuroimage, 44*(1), 175–181.

Penfield, W., & Boldrey, E. (1937). Somatic motor and sensory representations in the cerebral cortex of man as studied by electrical stimulation. *Brain*, *60*, 389–443.

Penfield, W., & Rasmussen, T. (1950). *The cerebral cortex of man: A clinical study of localization of function*. New York: MacMillan.

Perlman, M., & Clark, N. (2015). Learned vocal and breathing behavior in an enculturated gorilla. *Animal Cognition, 18*(5), 1165–1179. doi:10.1007/s10071-015-0889-6

Perry, D. W., Zatorre, R. J., Petrides, M., Alivisatos, B., Meyer, E., & Evans, a C. (1999). Localization of cerebral activity during simple singing. *Neuroreport, 10*(18), 3979–3984. doi:10.1097/00001756-199912160-00046

Petkov, C. I., & Jarvis, E. D. (2012). Birds, primates, and spoken language origins: Behavioral phenotypes and neurobiological substrates. *Frontiers in Evolutionary Neuroscience, 4*, 1–24. doi:10.3389/fnevo.2012.00012

Petrides, M., & Pandya, D. N. (2009). Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biology*, *7*(8). doi:10.1371/journal.pbio.1000170

Pfenning, A. R., Hara, E., Whitney, O., Rivas, M. V., Wang, R., Roulhac, P. L., … Jarvis, E. D. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science*, *346*(6215), 1256846–1256846. doi:10.1126/science.1256846

Pierce, J. D. (1985). A review of attempts to condition operantly alloprimate vocalizations. *Primates*, *26*(2), 202–213. doi:10.1007/BF02382019

Pisanski, K., Cartei, V., McGettigan, C., Raine, J., & Reby, D. (2016). Voice modulation: A window into the origins of human vocal control? *Trends in Cognitive Sciences, 20*, 304–318. doi:10.1016/j.tics.2016.01.002

Pisanski, K., Fraccaro, P. J., Tigue, C. C., O'Connor, J. J. M., Röder, S., Andrews, P. W., … Feinberg, D. R. (2014). Vocal indicators of body size in men and women: A meta-analysis. *Animal Behaviour*, *95*, 89–99. doi:10.1016/j.anbehav.2014.06.011

Poole, J. H., Tyack, P. L., Stoeger-Horwath, A. S., & Watwood, S. (2005). Elephants are capable of vocal learning. *Nature*, *434*, 455–456. doi:10.1029/2001GL014051

Porges, S. W. (2001). The polyvagal theory: Phylogenetic substrates of a social nervous system. *International Journal of Psychophysiology*, *42*(2), 123–146. doi:10.1016/S0167-8760(01)00162-3

Prades, J. M., Dubois, M. D., Dumollard, J. M., Tordella, L., Rigail, J., Timoshenko, A. P., & Peoc'h, M. (2012). Morphological and functional asymmetry of the human recurrent laryngeal nerve. *Surgical and Radiologic Anatomy*, *34*(10), 903–908. doi:10.1007/s00276-012-0999-7

Ralls, K., Fiorelli, P., & Gish, S. (1985). Vocalizations and vocal mimicry in captive harbor seals, Phoca vitulina. *Canadian Journal of Zoology*, *63*, 1050–1056.

Ramsay, S. C., Adams, L., Murphy, K., Corfield, D. R., Grootoonk, S., Bailey, D. L., … Guz, A. (1993). Regional cerebral blood flow during volitional expiration in man: A comparison with volitional inspiration. *Journal of Physiology*, *461*(1993), 85–101.

Redies, C. (1995). Cadherin expression in the developing vertebrate CNS: from neuromeres to brain nuclei and neural circuits. *Experimental Cell Research*, *220*, 243–256. doi:10.1006/excr.1995.1313

Redies, C. (2000). Cadherins in the central nervous system. *Progress in Neurobiology*, *61*(6), 611–648. doi:10.1016/S0301-0082(99)00070-2

Reichmuth, C., & Casey, C. (2014). Vocal learning in seals, sea lions, and walruses. *Current Opinion in Neurobiology, 28*, 66–71. doi:10.1016/j.conb.2014.06.011

Riecker, A., Ackermann, H., Wildgruber, D., Dogil, G., & Grodd, W. (2000). Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. *Neuroreport, 11*(9), 1997–2000. doi:10.1097/00001756-200006260-00038

Rilling, J. K., Glasser, M. F., Jbabdi, S., Andersson, J., & Preuss, T. M. (2012). Continuity, divergence, and the evolution of brain language pathways. *Frontiers in Evolutionary Neuroscience*, *3*, 1–6. doi:10.3389/fnevo.2011.00011

Rilling, J. K., Glasser, M. F., Preuss, T. M., Ma, X., Zhao, T., Hu, X., & Behrens, T. E. J. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nature Neuroscience*, *11*(4), 426–428. doi:10.1038/nn2072

Rödel, R. M. W., Laskawi, R., & Markus, H. (2003). Tongue representation in the lateral cortical motor region of the human brain as assessed by transcranial magnetic stimulation. *Annals of Otology, Rhinology and Laryngology, 112*, 71–76.

Rödel, R. M. W., Olthoff, A., Tergau, F., Simonyan, K., Kraemer, D., Markus, H., & Kruse, E. (2004). Human cortical motor representation of the larynx as assessed by transcranial magnetic stimulation (TMS). *Laryngoscope, 114*(5), 918–922. doi:10.1097/00005537-200405000-00026

Ross, E. D. (1981). The aprosodias: Functional-anatomic organization of the affective components of language in the right hemisphere. *Archives of Neurology, 38*(9), 561–569. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7271534

Ross, E. D., & Mesulam, M. (1979). Dominant language functions of the prosody and emotional gesturing right hemisphere ? *Archives of Neurology, 36*, 144–148.

Roubeau, B., Chevrie-Muller, C., & Saint Guily, J. (1997). Electromyographic activity of strap and cricothyroid muscles in pitch change. *Acta Otolaryngologica*, *117*(3), 459–464.

Rujescu, D., Meisenzahl, E. M., Krejcova, S., Giegling, I., Zetzsche, T., Reiser, M., … Finckh, U. (2007). Plexin B3 is genetically associated with verbal performance and white matter volume in human brain. *Molecular Psychiatry*, *12*(2), 190–194, doi:10.1038/sj.mp.4001903

Sanvito, S., Galimberti, F., & Miller, E. H. (2007). Observational evidences of vocal learning in southern elephant seals: A longitudinal study. *Ethology*, *113*(2), 137–146. doi:10.1111/j.1439-0310.2006.01306.x

Sapir, S., Campbell, C., & Larson, C. (1981). Effect of geniohyoid, cricothyroid and sternothyroid muscle stimulation on voice fundamental frequency of electrically elicited phonation in rhesus macaque. *Laryngoscope*, *91*(3), 457–468.

Schreiber, J., Riffert, T., Anwander, A., & Knösche, T. R. (2014). Plausibility tracking: A method to evaluate anatomical connectivity and microstructural properties along fiber pathways. *Neuroimage, 90*, 163–178. doi:10.1016/j.neuroimage.2014.01.002

Schröder, M. (2003). Experimental study of affect bursts. *Speech Communication, 40*, 99–116.

Schusterman, R. J., & Feinstein, S. H. (1965). Shaping and discriminative control of underwater click vocalizations in a California Sea Lion. *Science*, *150*(3704), 1743–1744. doi:10.1126/science.150.3704.1743

Seikel, J., King, D., & Drumwright, D. (2010). Anatomy of phonation. In *Anatomy & Physiology for Speech, Language, and Hearing* 4th ed. (pp. 165–222). Clifton Park, NY: Delmar, Cengage Learning.

Shipp, T., & Izdebski, K. (1975). Vocal frequency and vertical larynx positioning by singers. *Journal of the Acoustical Society of America*, *58*(5), 1104–1106.

Simonyan, K. (2014). The laryngeal motor cortex: Its organization and connectivity. *Current Opinion in Neurobiology*, *28*, 15–21.

Simonyan, K., & Horwitz, B. (2011). Laryngeal motor cortex and control of speech in humans. *The Neuroscientist*, *17*(2), 197–208. doi:10.1177/1073858410386727

Simonyan, K., & Jürgens, U. (2002). Cortico-cortical projections of the motorcortical larynx area in the rhesus monkey. *Brain Research*, *949*(1-2), 23–31.

Simonyan, K., & Jürgens, U. (2003). Efferent subcortical projections of the laryngeal motorcortex in the rhesus monkey. *Brain Research*, *974*(1-2), 43–59.

Simonyan, K., Ostuni, J., Ludlow, C. L., & Horwitz, B. (2009). Functional but not structural networks of the human laryngeal motor cortex show left hemispheric lateralization during syllable but not breathing production. *Journal of Neuroscience*, *29*(47), 14912–14923. doi:10.1523/JNEUROSCI.4897-09.2009

Simonyan, K., Saad, Z. S., Loucks, T. M. J., Poletto, C. J., & Ludlow, C. L. (2007). Functional neuroanatomy of human voluntary cough and sniff production. *Neuroimage, 37*(2), 401–409. doi:10.1016/j.neuroimage.2007.05.021

Sohrabji, F., Nordeen, E. J., & Nordeen, K. W. (1990). Selective impairment of song learning following lesions of a forebrain nucleus in the juvenile zebra finch. *Behavioral and Neural Biology*, *53*(1), 51–63. doi:10.1016/0163-1047(90)90797-A

Sporns, O. (2006). Small-world connectivity, motif composition, and complexity of fractal neuronal connections. *Biosystems*, *85*, 55–64.

Sporns, O., & Zwi, J. D. (2004). The small world of the cerebral cortex. *Neuroinformatics*, *2*, 145–162.

Steinecke, I., & Herzel, H. (1995). Bifurcation in an asymmetric vocal-fold model. *Journal of the Acoustical Society of America*, *97*, 1874–1884.

Stoeger, A. S., & Manger, P. (2014). Vocal learning in elephants: Neural bases and adaptive context. *Current Opinion in Neurobiology, 28*, 101–107. doi:10.1016/j.conb.2014.07.001

Stoeger, A. S., Mietchen, D., Oh, S., de Silva, S., Herbst, C. T., Kwon, S., & Fitch, W. T. (2012). An Asian elephant imitates human speech. *Current Biology*, *22*(22), 2144–2148. doi:10.1016/j.cub.2012.09.022

Takahashi, T., Fournier, a, Nakamura, F., Wang, L. H., Murakami, Y., Kalb, R. G., … Strittmatter, S. M. (1999). Plexin-neuropilin-1 complexes form functional semaphorin-3A receptors. *Cell*, *99*(1), 59–69.

Tamagnone, L., Artigiani, S., Chen, H., He, Z., Ming, G. I., Song, H., … Comoglio, P. M. (1999). Plexins are a large family of receptors for transmembrane, secreted, and GPI-anchored semaphorins in vertebrates. *Cell*, *99*(1), 71–80.

Thiebaut de Schotten, M., Dell'Acqua, F., Valabregue, R., & Catani, M. (2012). Monkey to human comparative anatomy of the frontal lobe association tracts. *Cortex*, *48*(1), 82–96. doi:10.1016/j.cortex.2011.10.001

Thumfart, W. F. (1988). Electrodiagnosis of laryngeal nerve disorders. *Ear, Nose, & Throat Journal*, *67*, 380–393.

Thumfart, W.F., Pototschnig, C., Zorowka, P., & Eckel, H. E. (1992). Electrophysiologic investigation of lower cranial nerve diseases by means of magnetically stimulated neuromyography of the larynx. *Annals of Otology, Rhinology & Laryngology*, *101*, 629–634.

Titze, I. R. (2008). Nonlinear source–filter coupling in phonation: Theory. *Journal of the Acoustical Society of America*, *123*(5), 2733. doi:10.1121/1.2832337

Titze, I. R., Luschei, E. S., & Hirano, M. (1989). Role of the thyroarytenoid muscle in regulation of fundamental frequency. *Journal of Voice*, *3*(3), 213–224. doi:10.1016/S0892-1997(89)80003-7

Titze, I. R., & Story, B. H. (2002). Rules for controlling low-dimensional vocal fold models with muscle activation. *Journal of the Acoustical Society of America*, *112*, 1064–1076. doi:10.1121/1.1496080

Townsend, S., & Zuberbuhler, K. (2009). Audience effects in chimpanzee copulation calls. *Communicative and Integrative Biology*, *2*(3), 282–284. doi:10.4161/cib.2.3.6796

Trauner, D. A., Ballantyne, A., Friedland, S., & Chase, C. (1996). Disorders of affective and linguistic prosody in children after early unilateral brain damage. *Annals of Neurology, 39*(3), 361–367. doi:10.1002/ana.410390313

Turkeltaub, P. E., Eden, G. F., Jones, K. M., & Zeffiro, T. A. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: Method and validation. *Neuroimage, 16*(3), 765–780. doi:10.1006/nimg.2002.1131

VanderHorst, V. G. J. M., Terasawa, E., & Ralston, H. J. (2001). Monosynaptic projections from the nucleus retroambiguus region to laryngeal motoneurons in the rhesus monkey. *Neuroscience*, *107*, 117–125. doi:10.1016/S0306-4522(01)00343-8

Vanderhorst, V. G. J. M., Terasawa, E., Ralston, H. J., & Holstege, G. (2000). Monosynaptic projections from the lateral periaqueductal gray to the nucleus retroambiguus in the rhesus monkey: Implications for vocalization and reproductive behavior. *Journal of Comparative Neurology, 424*, 251–268. doi:10.1002/1096-9861(20000821)424:2<251::AID-CNE5>3.0.CO;2-D

Vaneechoutte, M., Munro, S., & Verhaegen, M. (2011). Seafood, diving, song and speech. In M. Vaneechoutte, A. Kuliukas & M. Verhaegen (Eds.), *Was man more aquatic in the past?* (pp. 181-189). Bentham Science ebooks.

Vernes, S. C. (2016). What bats have to say about speech and language. *Psychonomic Bulletin & Review*, 1–7. doi:10.3758/s13423-016-1060-3

Vilkman, E., Sonninen, A., Hurme, P., & Körkkö, P. (1996). External laryngeal frame function in voice production revisited: A review. *Journal of Voice*, *10*(1), 78–92.

Vogt, O. (1910). Die myeloarchitektonische Felderung des menschlichen Stirnhirns. *Journal Für Psychologie Und Neurologie*, 15, 221–232.

Wahl, M., Lauterbach-Soon, B., Hattingen, E., Jung, P., Singer, O., Volz, S., … Ziemann, U. (2007). Human motor corpus callosum: Topography, somatotopy, and link between microstructure and function. *Journal of Neuroscience*, *27*(45), 12132–12138. doi:10.1523/JNEUROSCI.2320-07.2007

Walker, S. F. (1994). The possible role of asymmetric laryngeal innervation in language lateralization: Points for and against. *Brain and Language*, *46*, 482–489.

Walker, A. E., & Green, H. D. (1938). Electrical excitability of the motor face area: A comparative study in primates. *Journal of Neurophysiology*, *1*, 152–165.

Wattendorf, E., Westermann, B., Fiedler, K., Kaza, E., Lotze, M., & Celio, M. R. (2013). Exploration of the neural correlates of ticklish laughter by functional magnetic resonance imaging. *Cerebral Cortex*, *23*, 1280–1289. doi:10.1093/cercor/bhs0

Ward, B. C., Nordeen, E. J., & Nordeen, K. W. (1998). Individual variation in neuron number predicts differences in the propensity for avian vocal imitation. *Proceedings of the National Academy of Sciences*, *95*, 1277–1282.

Wich, S. A., Krützen, M., Lameira, A. R., Nater, A., Arora, N., Bastian, M. L., … van Schaik, C. P. (2012). Call cultures in orang-utans? *PLoS ONE, 7*(5), 1–9. doi:10.1371/journal.pone.0036180

Wich, S. A, Swartz, K. B., Hardus, M. E., Lameira, A. R., Stromberg, E., & Shumaker, R. W. (2009). A case of spontaneous acquisition of a human sound by an orangutan. *Primates*, *50*(1), 56–64. doi:10.1007/s10329-008-0117-y

Wild, J. M. (1993). Descending projections of the songbird nucleus robustus archistriatalis. *Journal of Comparative Neurology, 338*(2), 225–241. doi:10.1002/cne.903380207

Wild, J. M., Kubke, M. F., & Mooney, R. (2009). Avian nucleus retroambigualis: Cell types and projections to other respiratory-vocal nuclei in the brain of the Zebra Finch (Taeniopygia guttata). *Journal of Computational Neuroscience, 512*(6), 768–783. doi:10.1016/j.surg.2006.10.010.Use

Zatorre, R. J., & Baum, S. R. (2012). Musical melody and speech intonation: Singing a different tune. *PLoS Biology, 10*(7), e1001372. doi:10.1371/journal.pbio.1001372

Legends

Figure 1: Superior view of the larynx. The arrows represent a simplification of the two dimensions of movement that most strongly influence vocalization. First, rotation of the arytenoid cartilages causes adduction or abduction of the vocal folds to control the onset or offset, respectively, of voicing. Second, the thyroid cartilage can rock forward or backward to affect the tension of the vocal folds so as to modulate vocal pitch. The drawing is modified from Gray (1918).

Figure 2. The dual structure of the larynx motor cortex in humans. The ventral part of the larynx motor cortex (LMC) in the Rolandic operculum is proposed to be the human homologue of the non-human primate LMC. The dorsal LMC in the facial region of the motor cortex is proposed to be a novel human area. Schematized locations of the dorsal (orange) and ventral (green) LMC are shown at the extremes of the orofacial representation of the primary motor cortex. The region colored in purple on the anatomical brain is the primary motor cortex in the precentral gyrus.

Figure 3. A model of efferent pathways relevant for vocalization. The descending corticobulbar pathways relevant for vocalization are described diagrammatically. Solid lines represent known pathways (see Supplementary Figure 1 for supporting references). An efferent pathway from the LMCs to respiratory motor neurons is conspicuously absent, despite the involvement of the dorsal LMC in expiration. The dashed line represents an hypothesized, but as-yet-unobserved, efferent pathway (either direct or indirect) from the LMC to the nucleus retroambiguus, which contains respiratory motor neurons. Abbreviation: ACC, anterior cingulate cortex; dLMC, dorsal larynx motor cortex; PAG: periaqueductal grey; vLMC, ventral larynx motor cortex.

Figure 4. The "single vocal system" model. The model proposes that, although language, emotion, and music, are different systems for communicating meaning, they feed into a common sensorimotor vocal system to produce speech, emotional vocalizations, and song as their respective vocal outputs.

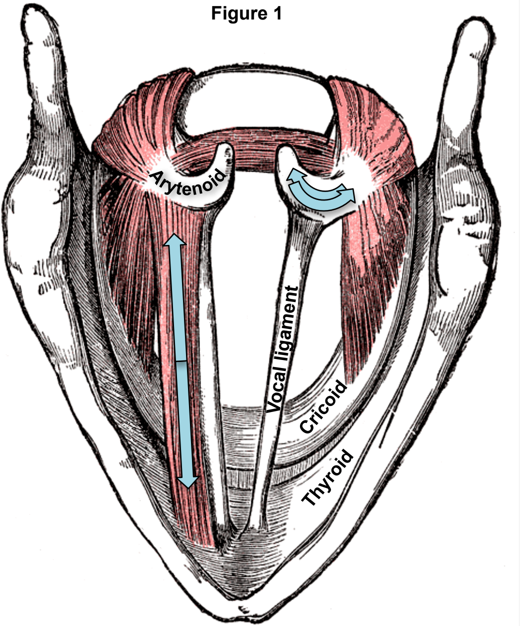
Figure 5. Evolutionary scenarios for the emergence of the human larynx motor cortex. Two evolutionary models are presented. In both models, a posterior migration of the larynx motor cortex (LMC) is proposed to occur from the monkey to the ape positions by means of relocation. Hence, in both models, the ventral LMC of humans is seen as the homologue of the ape LMC that migrated further posteriorly along the inferior part of the frontal lobe as a second occurrence of relocation. For ease of comparison, the approximate positions of the monkey and ape LMCs are shown on a human brain, rather than on three species-specific brains. The left panel depicts the "duplication + migration model", according to which the dorsal LMC arose by a duplication of motor regions related to vocalization – such as the ventral LMC and/or the trunk motor cortex controlling respiration – followed by a migration into the orofacial region of the motor cortex. The right panel depicts the "local duplication model", according to which the dorsal LMC evolved by duplication of an adjacent, though non-vocal, region of the motor cortex. The diamond-shaped region colored pale orange signifies the potential source-regions for such a duplication, including orofacial regions ventral to the dorsal LMC and hand-controlling regions dorsal to it. The region colored in purple on the anatomical brain is the primary motor cortex of the precentral gyrus. Yellow arrows signify migrations that result in the relocation of a brain area between species. Red arrows signify migrations (either distant or local) that occur following brain-area duplication, resulting in the establishment of a novel brain area.

Figure 6. Comparative analysis of vocalizing and the vocal brain in primates. This summary provides a phylogenetic view of the behavioral capacities and neural structures related to vocalization and vocal learning. The left side sketches a rough phylogeny of the groups of species for which neuroscientific data are available for comparison, namely monkeys (macaque and squirrel monkey), non-human great apes (chimpanzees, gorillas, and orangutans), and humans. The right side highlights changes throughout the vocal-motor system during primate evolution. Colors correspond to the representations of the same brain areas in Figures 2, 3, 5, and 7. See Supplementary Figure 2 for supporting references. Abbreviation: mya, million years ago.

Figure 7. An evolutionary model for the joint emergence of vocal-motor control and vocal learning in humans through brain pathway duplication. The left panel shows the ancestral state of a non-vocal audio-motor system, with arcuate fasciculus (AF) connections between auditory areas and the inferior frontal gyrus (IFG), which projects to the ancestral, non-vocal larynx motor cortex (LMC). The right panel presents an hypothesis for the joint emergence of voluntary control of vocalization and vocal production learning via a process of *brain pathway duplication*, where expanded and potentially duplicated brain regions and pathways are highlighted with purple shading, in comparison to the ancestral pathway. An "audiovocal hub" in the inferior frontal gyrus (IFG) is proposed to emerge from an expansion of the IFG during human brain evolution. According to this model, the IFG hub emerged as an integration zone by: acquiring innervation from newly-evolved fibers during AF expansion, developing a novel projection to the newly-evolved dorsal LMC, and interfacing with the vocal part of the basal ganglia to mediate vocal imitation (not shown). As also shown in Figure 5, the dorsal LMC is proposed to be a duplication

and integration of components from the ventral LMC and expiratory motor cortex. The ventral LMC has relocated to a more posterior position in the inferior frontal lobe.

Figure 1



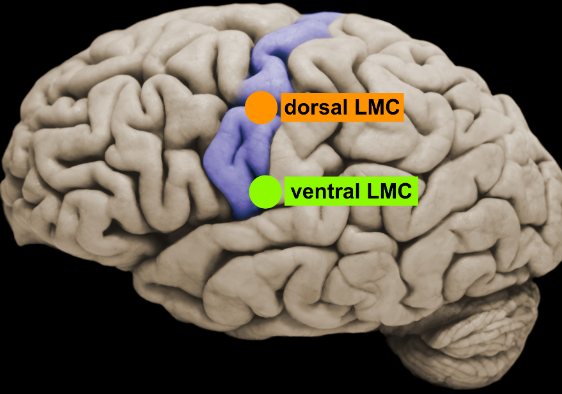
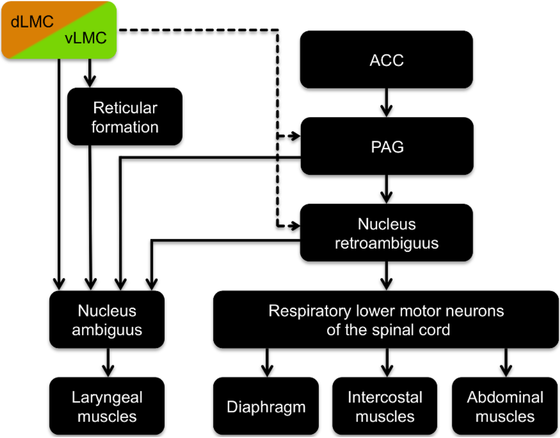


Figure 2


dLMC
vLMC
ACC
Reticular formation
PAG
Nucleus retroambiguus
Nucleus ambiguus
Respiratory lower motor neurons of the spinal cord
Laryngeal muscles
Diaphragm
Intercostal muscles
Abdominal muscles

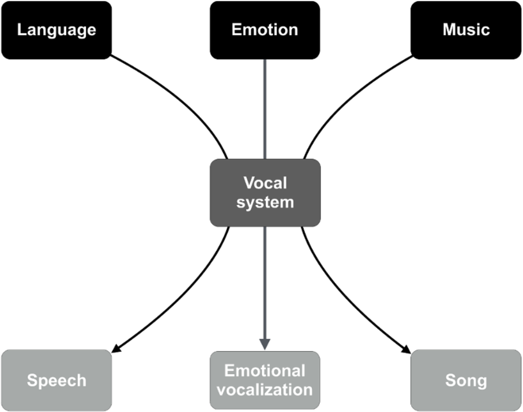
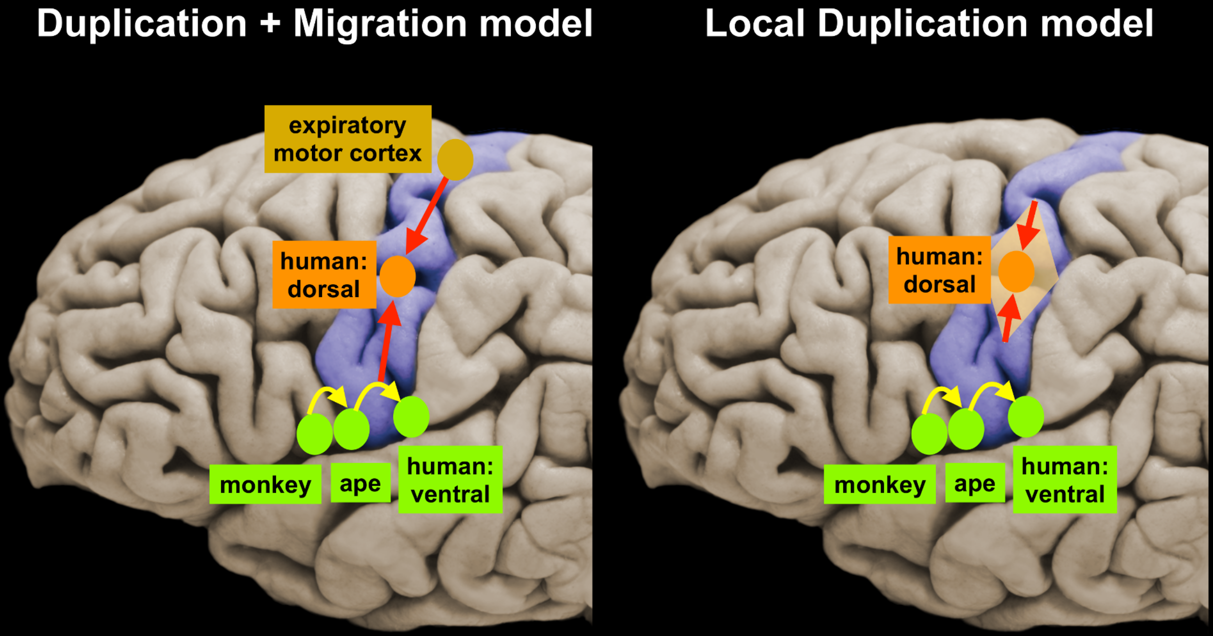

Language
Emotion
Music
Vocal system
Speech
Emotional vocalization
Song

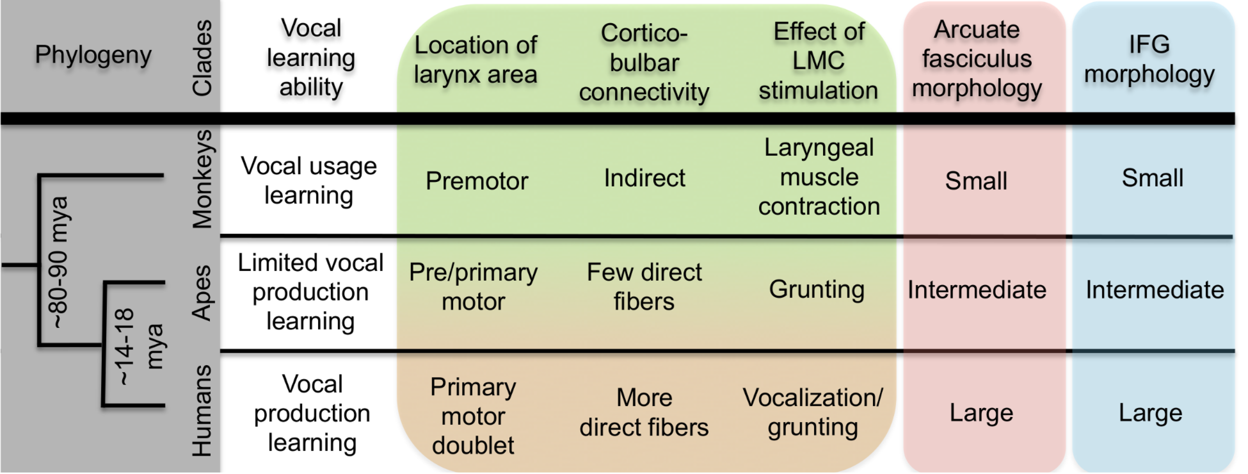
Figure 4



Duplication + Migration model

Local Duplication model

Figure 5



Phylogeny	Clades	Vocal learning ability	Location of larynx area	Cortico-bulbar connectivity	Effect of LMC stimulation	Arcuate fasciculus morphology	IFG morphology
	Monkeys	Vocal usage learning	Premotor	Indirect	Laryngeal muscle contraction	Small	Small
	Apes	Limited vocal production learning	Pre/primary motor	Few direct fibers	Grunting	Intermediate	Intermediate
	Humans	Vocal production learning	Primary motor doublet	More direct fibers	Vocalization/ grunting	Large	Large

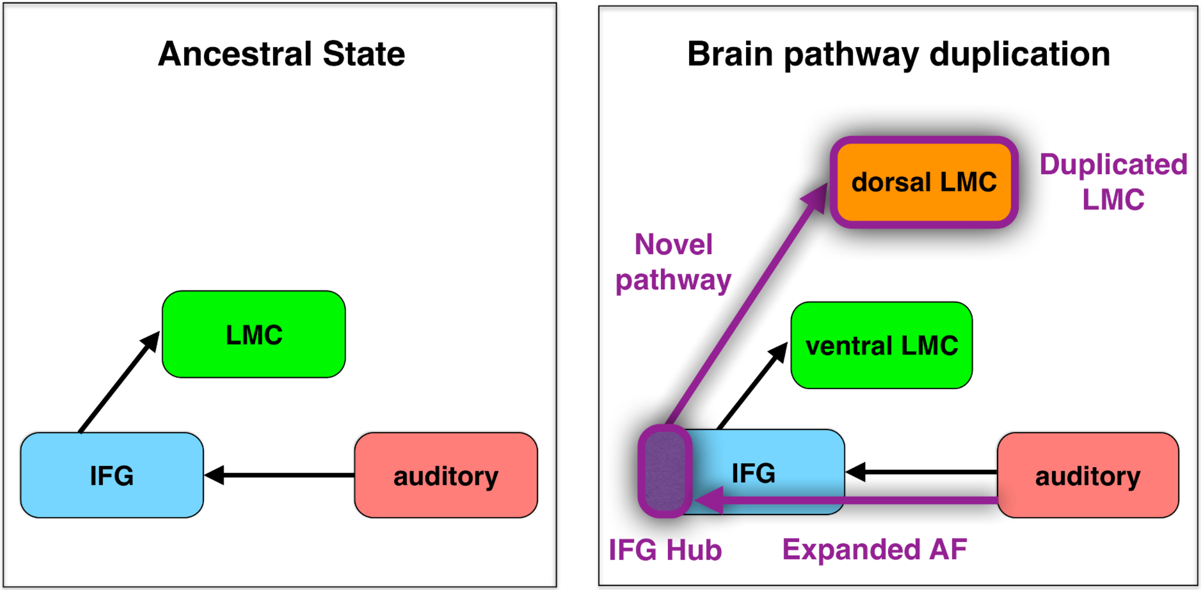
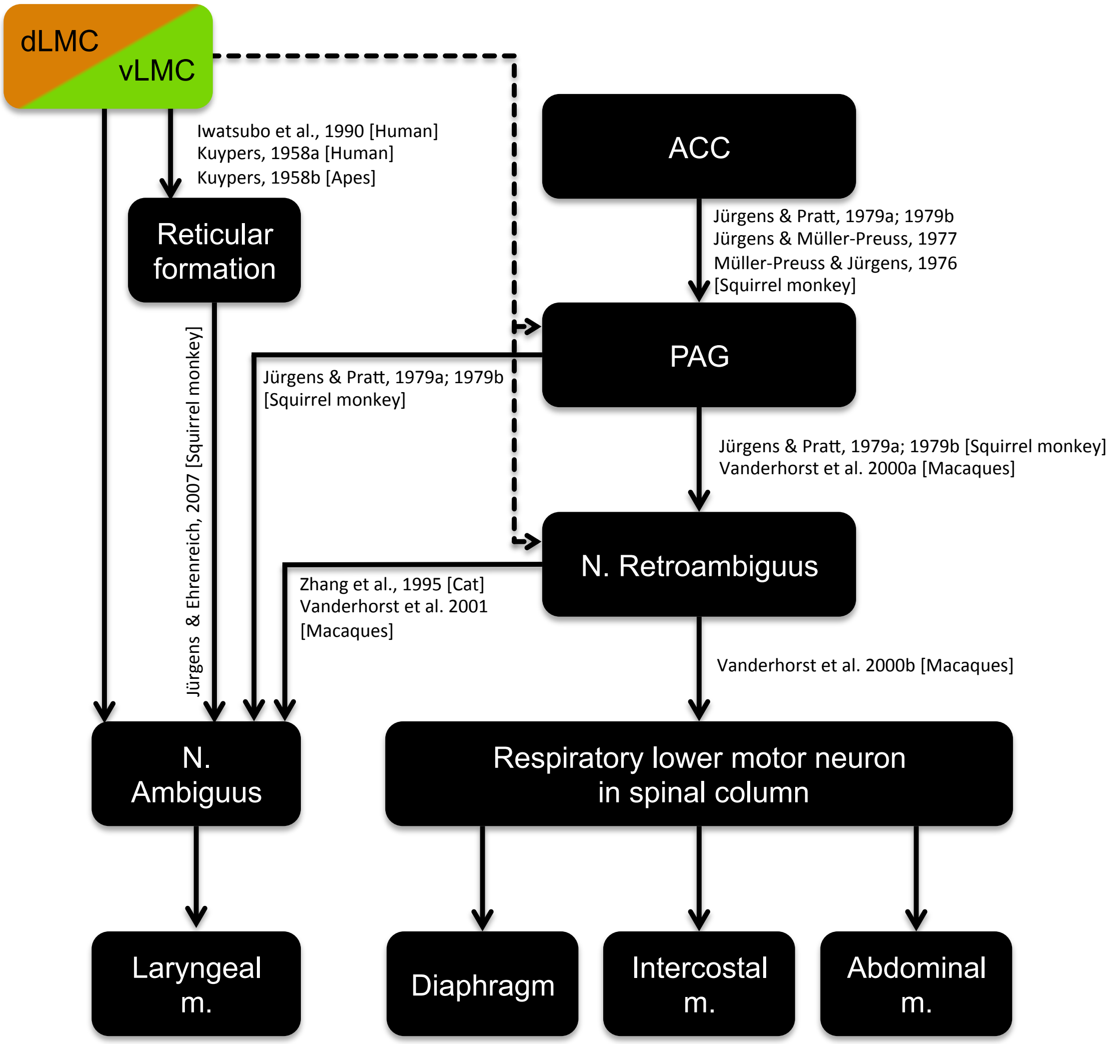

Ancestral State
LMC
IFG
auditory
Brain pathway duplication
dorsal LMC
Duplicated LMC
Novel pathway
ventral LMC
IFG
auditory
IFG Hub
Expanded AF

Figure 7



Supplementary Figure 1: Enlargement of Figure 2 with selected references supporting pathways that have been described by neurophysiological tracer studies. In each case the study species in noted in parentheses.

Figure References

Iwatsubo, T., Kuzuhara, S., & Kanemitsu, A. (1990). Corticofugal projections to the motor nuclei of the brainstem and spinal cord in humans. *Neurology*, *40*, 309–312. Retrieved from http://www.neurology.org/content/40/2/309.short

Jürgens, U., & Ehrenreich, L. (2007). The descending motorcortical pathway to the laryngeal motoneurons in the squirrel monkey. *Brain Research*, *1148*, 90–95. doi:10.1016/j.brainres.2007.02.020

Jürgens, U., & Müller-Preuss, P. (1977). Convergent projections of different limbic vocalization areas in the squirrel monkey. *Experimental Brain Research, 29*(1), 75–83. doi:10.1007/BF00236876

Jurgens, U., & Pratt, R. (1979). Role of the periaqueductal grey in vocal expression of emotion. *Brain Research*, *167*, 367–378.

Jürgens, U., & Pratt, R. (1979). The cingular vocalization pathway in the squirrel monkey. *Experimental Brain Research*, *510*, 499–510. Retrieved from http://link.springer.com/article/10.1007/BF00239145

Kuypers, H. G. J. M. (1958). Corticobulbar Connexions To the Pons and Lower Brain-Stem in Man. *Brain*, *81*(3), 364–388. doi:10.1093/brain/81.3.364

Kuypers, H. G. J. M. (1958). Some projections from the peri-central cortex to the pons and lower brain stem in monkey and chimpanzee. *The Journal of Comparative Neurology*, *110*, 221–255.

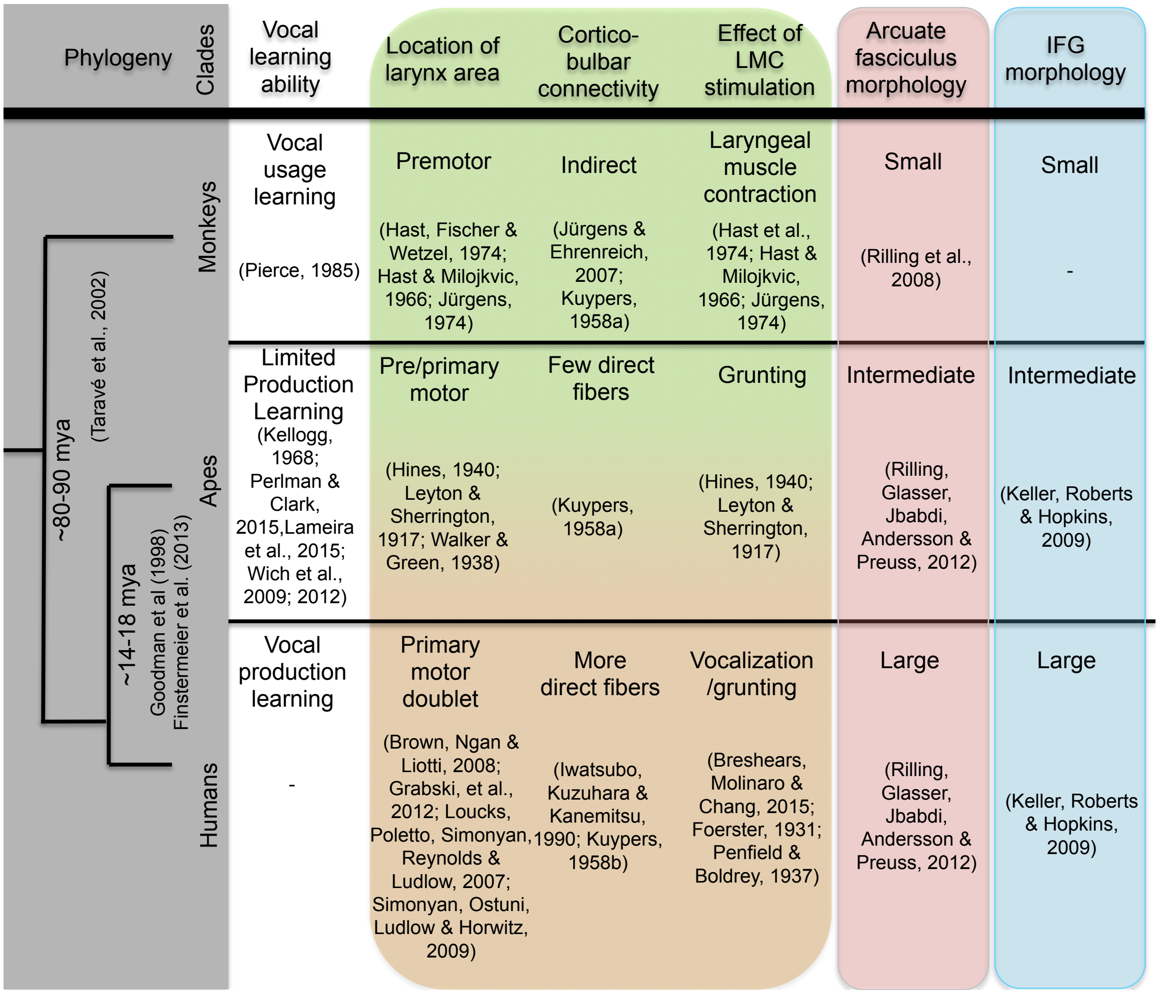
Müller-Preuss, P., & Jürgens, U. W. E. (1976). Projections from the "cingular" vocalization area in the squirrel monkey. *Brain Research, 103*, 29–43.

VanderHorst, V. G. J. M., Terasawa, E., & Ralston, H. J. (2001). Monosynaptic projections from the nucleus retroambiguus region to laryngeal motoneurons in the rhesus monkey. *Neuroscience*, *107*, 117–125. doi:10.1016/S0306-4522(01)00343-8

Vanderhorst, V. G. J. M., Terasawa, E., Ralston, H. J., & Holstege, G. (2000a). Monosynaptic projections from the lateral periaqueductal gray to the nucleus retroambiguus in the rhesus monkey: Implications for vocalization and reproductive behavior. *Journal of Comparative Neurology*, *424*, 251–268. doi:10.1002/1096-9861(20000821)424:2<251::AID-CNE5>3.0.CO;2-D

Vanderhorst, V. G. J. M., Terasawa, E., Ralston, H. J., & Holstege, G. (2000b). Monosynaptic projections from the nucleus retroambiguus to motoneurons supplying the abdominal wall, axial, hindlimb, and pelvic floor muscles in the female rhesus monkey. *Journal of Comparative Neurology*, *424*, 233–250. doi:10.1002/1096-9861(20000821)424:2<233::AID-CNE4>3.0.CO;2-C

Zhang, S. P., Bandler, R., & Davis, P. J. (1995). Brain stem integration of vocalization: Role of the nucleus retroambigualis. *Journal of Neurophysiology*, *74*(6), 2500–2512.



Clades	Vocal learning ability	Location of larynx area	Cortico-bulbar connectivity	Effect of LMC stimulation	Arcuate fasciculus morphology	IFG morphology
Monkeys	Vocal usage learning (Pierce, 1985)	Premotor (Hast, Fischer & Wetzel, 1974; Hast & Milojkvic, 1966; Jürgens, 1974)	Indirect (Jürgens & Ehrenreich, 2007; Kuypers, 1958a)	Laryngeal muscle contraction (Hast et al., 1974; Hast & Milojkvic, 1966; Jürgens, 1974)	Small (Rilling et al., 2008)	Small -
Apes	Limited Production Learning (Kellogg, 1968; Perlman & Clark, 2015,Lameira et al., 2015; Wich et al., 2009; 2012)	Pre/primary motor (Hines, 1940; Leyton & Sherrington, 1917; Walker & Green, 1938)	Few direct fibers (Kuypers, 1958a)	Grunting (Hines, 1940; Leyton & Sherrington, 1917)	Intermediate (Rilling, Glasser, Jbabdi, Andersson & Preuss, 2012)	Intermediate (Keller, Roberts & Hopkins, 2009)
Humans	Vocal production learning -	Primary motor doublet (Brown, Ngan & Liotti, 2008; Grabski, et al., 2012; Loucks, Poletto, Simonyan, Reynolds & Ludlow, 2007; Simonyan, Ostuni, Ludlow & Horwitz, 2009)	More direct fibers (Iwatsubo, Kuzuhara & Kanemitsu, 1990; Kuypers, 1958b)	Vocalization /grunting (Breshears, Molinaro & Chang, 2015; Foerster, 1931; Penfield & Boldrey, 1937)	Large (Rilling, Glasser, Jbabdi, Andersson & Preuss, 2012)	Large (Keller, Roberts & Hopkins, 2009)

Supplementary Figure 2: Enlargement of Figure 5 with supporting references.

Figure References

Bergman, T. J. (2013). Speech-like vocalized lip-smacking in geladas. *Current Biology*, *23*(7), R268–R2689. doi:10.1016/j.cub.2013.02.038

Breshears, J. D., Molinaro, A. M., & Chang, E. F. (2015). A probabilistic map of the human ventral sensorimotor cortex using electrical stimulation. *Journal of Neurosurgery*. doi:10.3171/2014.11.JNS14889.DISCLOSURE

Brown, S., Ngan, E., & Liotti, M. (2008). A larynx area in the human motor cortex. *Cerebral Cortex*, *18*(4), 837–845. doi: 10.1093/cercor/bhm131

Finstermeier, K., Zinner, D., Brameier, M., Meyer, M., Kreuz, E., Hofreiter, M., & Roos, C. (2013). A Mitogenomic Phylogeny of Living Primates. *PLoS ONE*, *8*(7), 1–10. http://doi.org/10.1371/journal.pone.0069504

Foerster, O. (1931). The cerebral cortex in man. *Lancet*, *218*(5632), 309–312.

Goodman, M., Porter, C. a, Czelusniak, J., Page, S. L., Schneider, H., Shoshani, J., … Groves, C. P. (1998). Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. *Molecular Phylogenetics and Evolution*, *9*(3), 585–598. doi.org/10.1006/mpev.1998.0495

Grabski, K., Lamalle, L., Vilain, C., Schwartz, J.-L., Vallée, N., Tropres, I., … Sato, M. (2012). Functional MRI assessment of orofacial articulators: Neural correlates of lip, jaw, larynx, and tongue movements. *Human Brain Mapping*, *33*(10), 2306–2321. doi:10.1002/hbm.21363

Hasegawa, M., Kishino, H., & Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, *22*(2), 160–174. doi:10.1007/BF02101694

Hast, M. H., Fischer, J. M., & Wetzel, A. B. (1974). Cortical motor representation of the laryngeal muscles in macaca mulatta. *Brain*, *73*, 229–240.

Hast, M. H., & Milojkvic, R. (1966). The response of the vocal folds to electrical stimulation of the inferior frontal cortex of the squirrel monkey. *Acta Oto-Laryngologica*, *61*, 196–204. doi:10.3109/00016486609127056

Hayes, C. (1952). *The ape on our house*. London: Victor Gollancz LTD.

Hines, M. (1940). Movements elicited from precentral gyrus of adult chimpanzees by stimulation with sine wave currents. *Journal of Neurophysiology*, *3*, 442–466.

Hopkins, W. D., Taglialatela, J., & Leavens, D. a. (2007). Chimpanzees Differentially Produce Novel Vocalizations to Capture the Attention of a Human. *Animal Behaviour*, *73*(2), 281–286. doi:10.1016/j.anbehav.2006.08.004

Iwatsubo, T., Kuzuhara, S., & Kanemitsu, A. (1990). Corticofugal projections to the motor nuclei of the brainstem and spinal cord in humans. *Neurology*, *40*, 309–312.

Jürgens, U. (1974). On the elicitability of vocalization from the cortical larynx area. *Brain Research*, *81*(3), 564–566.

Jürgens, U., & Ehrenreich, L. (2007). The descending motorcortical pathway to the laryngeal motoneurons in the squirrel monkey. *Brain Research*, *1148*, 90–95. doi:10.1016/j.brainres.2007.02.020

Keller, S. S., Roberts, N., & Hopkins, W. (2009). A comparative magnetic resonance imaging study of the anatomy, variability, and asymmetry of Broca's area in the human and chimpanzee brain. *The Journal of Neuroscience*, *29*(46), 14607–14616. doi:10.1523/JNEUROSCI.2892-09.2009

Kellogg, W. N. (1968). Communication and Language in the Home-Raised Chimpanzee. *Science*. doi:10.1126/science.162.3852.423

Kuypers, H. G. J. M. (1958). Corticobulbar Connexions To the Pons and Lower Brain-Stem in Man. *Brain*, *81*(3), 364–388. doi:10.1093/brain/81.3.364

Kuypers, H. G. J. M. (1958). Some projections from the peri-central cortex to the pons and lower brain stem in monkey and chimpanzee. *The Journal of Comparative Neurology*, *110*, 221–255.

Lameira, A. R., Hardus, M. E., Bartlett, A. M., Shumaker, R. W., Wich, S. a., & Menken, S. B. J. (2015). Speech-Like Rhythm in a Voiced and Voiceless Orangutan Call. *PLoS ONE*, *10*(1), e116136. doi:10.1371/journal.pone.0116136

Leyton, S., & Sherrington, C. (1917). Observations on the excitable cortex of the chimpanzee, organ-utan, and gorilla. *Experimental Physiology*, *11*(2), 135–222.

Loucks, T. M. J., Poletto, C. J., Simonyan, K., Reynolds, C. L., & Ludlow, C. L. (2007). Human brain activation during phonation and exhalation: Common volitional control for two upper airway functions. *NeuroImage*, *36*(1), 131–143. doi:10.1016/j.neuroimage.2007.01.049

Penfield, W., & Boldrey, E. (1937). Somatic motor and sensory representations in the cerebral cortex of man as studied by electrical stimulation. *Brain*, *60*, 389–443.

Perlman, M., & Clark, N. (2015). Learned vocal and breathing behavior in an enculturated gorilla. *Animal Cognition, 18*(5), 1165–1179. doi:10.1007/s10071-015-0889-6

Pierce, J. D. (1985). A review of attempts to condition operantly alloprimate vocalizations. *Primates*, *26*(2), 202–213. doi:10.1007/BF02382019

Rilling, J. K., Glasser, M. F., Jbabdi, S., Andersson, J., & Preuss, T. M. (2012). Continuity, divergence, and the evolution of brain language pathways. *Frontiers in Evolutionary Neuroscience*, *3*(January), 1–6. doi:10.3389/fnevo.2011.00011

Rilling, J. K., Glasser, M. F., Preuss, T. M., Ma, X., Zhao, T., Hu, X., & Behrens, T. E. J. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nature Neuroscience*, *11*(4), 426–428. doi:10.1038/nn2072

Simonyan, K., Ostuni, J., Ludlow, C. L., & Horwitz, B. (2009). Functional but not structural networks of the human laryngeal motor cortex show left hemispheric lateralization during syllable but not breathing production. *The Journal of Neuroscience*, *29*(47), 14912–14923. doi:10.1523/JNEUROSCI.4897-09.2009

Tavaré, S., Marshall, C. R., Will, O., Soligo, C., & Martin, R. D. (2002). Using the fossil record to estimate the age of the last common ancestor of extant primates. *Nature*, *416*(6882), 726–729. doi:10.1038/416726a

Walker, A. E., & Green, H. D. (1938). Electrical excitability of the motor face area: A comparative study in primates. *Journal of Neurophysiology*, *1*, 152–165.

Wich, S. A., Krützen, M., Lameira, A. R., Nater, A., Arora, N., Bastian, M. L., … van Schaik, C. P. (2012). Call cultures in orang-utans? *PLoS ONE, 7*(5), 1–9. doi:10.1371/journal.pone.0036180

Wich, S. a, Swartz, K. B., Hardus, M. E., Lameira, A. R., Stromberg, E., & Shumaker, R. W. (2009). A case of spontaneous acquisition of a human sound by an orangutan. *Primates*, *50*(1), 56–64. doi:10.1007/s10329-008-0117-y