

# UCSF

## UC San Francisco Previously Published Works

### Title

New Developments in Understanding the Complexity of Human Speech Production

### Permalink

<https://escholarship.org/uc/item/6ss8g8fk>

### Journal

Journal of Neuroscience, 36(45)

### ISSN

0270-6474

### Authors

Simonyan, Kristina  
Ackermann, Hermann  
Chang, Edward F  
et al.

### Publication Date

2016-11-09


### DOI

10.1523/jneurosci.2424-16.2016

Peer reviewed

## Symposium

# New Developments in Understanding the Complexity of Human Speech Production

 Kristina Simonyan,<sup>1</sup> Hermann Ackermann,<sup>2</sup> Edward F. Chang,<sup>3</sup> and Jeremy D. Greenlee<sup>4</sup>

<sup>1</sup>Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, New York 10029, <sup>2</sup>Center of Neurology, Hertie Institute for Clinical Brain Research, University of Tuebingen, Tuebingen 72076, Germany, <sup>3</sup>Department of Neurosurgery, University of California, San Francisco, San Francisco, California 94143, and <sup>4</sup>Department of Neurosurgery, University of Iowa, Iowa City, Iowa 52242

Speech is one of the most unique features of human communication. Our ability to articulate our thoughts by means of speech production depends critically on the integrity of the motor cortex. Long thought to be a low-order brain region, exciting work in the past years is overturning this notion. Here, we highlight some of major experimental advances in speech motor control research and discuss the emerging findings about the complexity of speech motocortical organization and its large-scale networks. This review summarizes the talks presented at a symposium at the Annual Meeting of the Society of Neuroscience; it does not represent a comprehensive review of contemporary literature in the broader field of speech motor control.

**Key words:** ECoG; motor cortex; neuroimaging; speech production

## Introduction

The power of speaking cannot be underestimated as it allows us to express who we are, our intentions, hopes, and beliefs. As a result, the neural mechanisms of voice, speech, and language control have been a topic of intense investigations for centuries. However, with the major focus on perceptual and cognitive aspects of speech and language processing, little attention has been given to the motocortical control of speech production. This is due, in part, to the continuous major technical challenges in this field associated with the absence of animal models of real-life speaking and a limited range of invasive studies that can be performed in humans to assess the neural bases of this complex behavior.

Voice production is controlled by a hierarchically organized, bottom-up neural system that extends from the control of innate vocalizations (lower brainstem and spinal cord) to the control of vocalization initiation, motivation, and expression of voluntary emotional vocalization (periaqueductal gray, limbic structures, and cingulate cortex) to voluntary vocal motor control (laryngeal/orofacial motor cortex with its input and output structures) (Jürgens, 2002; Simonyan and Horwitz, 2011; Ackermann et al., 2014) (Fig. 1). The human ability to gradually acquire and produce more complex vocalizations, from basic nonverbal vocal

reactions to voluntary speech production, is based on the maturation and modulation of this system along the development. Recent evidence suggests that nonhuman primates and great apes may also be able to modulate their nonverbal vocalizations during development (Takahashi et al., 2015) and in adulthood (Lameira et al., 2016). However, the highest level of voluntary motor control of their articulate “speech” appears to lack significant vocal antecedents within their lineages and the characteristic complexity of human speech (Jürgens, 2002; Simonyan and Horwitz, 2011; Ackermann et al., 2014). For example, all attempts to teach great apes real-life spoken language have failed, although these species have highly mobile lips and tongue, often exceeding the respective motor capabilities in humans. Conceivably, nonhuman primates are, by and large, unable to decouple the laryngeal sound source from genetically preprogrammed and phylogenetically adapted vocal “fixed action” patterns (Winter et al., 1973; Kirzinger and Jürgens, 1982; Gemba, 2002; Jürgens, 2002; Arbib et al., 2008; Hage, 2010; Simonyan and Horwitz, 2011; Hage et al., 2013; Ackermann et al., 2014), precluding the production of a large variety of complex syllable-like utterances. Thus, although various animal models, including nonhuman primates, can be successfully used for examining the neural bases of other aspects of speech control (e.g., nonverbal vocalizations, acoustic voice perception and processing), humans remain the only species that can be studied in methodologically demanding experiments to assess motocortical control of voluntary speech production.

To that end, recent advances in mapping human brain organization have invigorated the interest in speech motor control. Combined knowledge derived from noninvasive and limited invasive studies of the central control of speech production is critically important as these methodologies are highly complementary and, at the same time, confirmatory of each other’s findings. An array of high-resolution noninvasive neuroimaging techniques has been

Received July 29, 2016; revised Sept. 20, 2016; accepted Sept. 21, 2016.

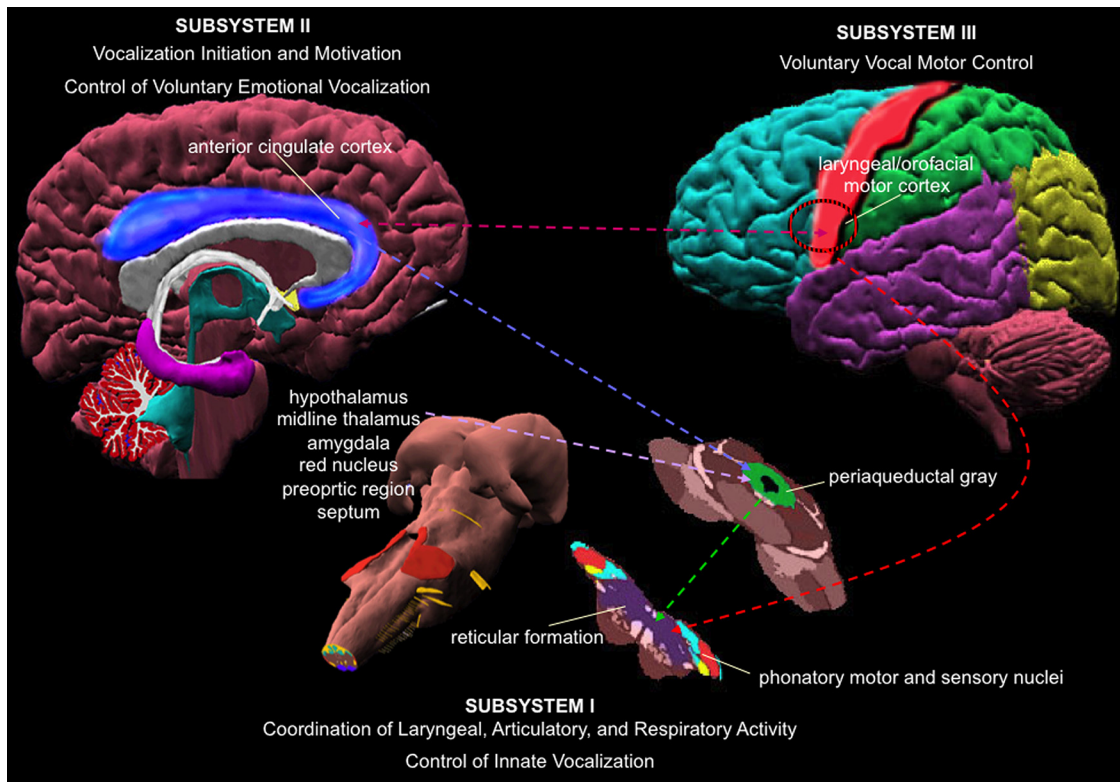
K.S. was supported by the National Institute on Deafness and Other Communication Disorders and National Institute of Neurological Disorders and Stroke, National Institutes of Health R01DC011805, R01DC0123545, and R01NS088160. E.F.C. was supported by the National Institute on Deafness and Other Communication Disorders, National Institutes of Health R01DC012379 and Office of the Director, National Institutes of Health DP2-OD00862. J.D.G. was supported by the National Institute on Deafness and Other Communication Disorders, National Institutes of Health R01DC015260.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Kristina Simonyan, Department of Neurology, One Gustave L. Levy Place, Box 1137, Icahn School of Medicine at Mount Sinai, New York, NY 10029. E-mail: kristina.simonyan@mssm.edu.

DOI:10.1523/JNEUROSCI.2424-16.2016

Copyright © 2016 the authors 0270-6474/16/3611440-09\$15.00/0



**Figure 1.** Hierarchical organization of the dual pathway of central voice control. The lowest level (subsystem I) is represented by the sensorimotor phonatory nuclei in brainstem and spinal cord, which control laryngeal, articular, and respiratory muscles during production of innate vocalizations. The higher level within this system (subsystem II) is represented by the periaqueductal gray, cingulate cortex, and limbic input structures that control vocalization initiation and motivation as well voluntary emotional vocalizations. The highest level (subsystem III) is represented by the laryngeal/orofacial motor cortex in the vSMC with its input and output regions that are responsible for voluntary motor control of speech production. Dotted lines indicate direct connections between different regions within the voice-controlling system. Data from Simonyan and Horwitz (2011).

successfully used in healthy and diseased individuals to examine different aspects of speech production. At the same time, human patient volunteers undergoing neurosurgical treatment to remove brain tumor or epileptogenic foci during awake craniotomy as well as to temporarily implant electrode arrays for localization and modulation of pathologic states provide a unique opportunity to evaluate and refine our understanding of neural mechanisms underlying speech motor control. In this regard, electrocorticography (ECoG) studies provide an unprecedented combination of temporal (in milliseconds) and spatial (in millimeters) resolution along with a frequency bandwidth (up to hundreds of Hertz) that noninvasive imaging methods are not able to match. However, a potential caveat of these recordings is that they are not performed on entirely neurologically healthy brains, although electrophysiological data argue that recordings from unaffected brain regions do reflect normal brain function (Lachaux et al., 2012).

In this brief review, we highlight the detailed organization of the ventral sensorimotor cortex (vSMC) for speech production; discuss unique recordings of speech motocortical activity that identified the specialized function of the speech motor cortex; examine the organization of large-scale neural networks controlling speech production; and discuss the role of subcortical structures, such as the basal ganglia and cerebellum, in driving speech preparation, execution, and motor skill acquisition. We make an attempt to shift the persisting view of the speech motor cortex as a low-order unimodal brain region (Callan et al., 2006; Hickok and Poeppel, 2007; Hickok et al., 2011; Poeppel et al., 2012; Tankus et al., 2014; Guenther and Hickok, 2015; Kawai et al., 2015) by discussing its organizational diversity and operational heterogeneity.

**The organization of the vSMC for speech control**

Speech production is one of the most complex and rapid motor behaviors. It depends on the precise coordination of >100 laryngeal, orofacial, and respiratory muscles whose neural representations are located within the vSMC. Injury to this brain area causes impairment of movement of muscles controlling speech production (dysarthria), whereas bilateral damage to vSMC leads to inability to produce voluntary vocalizations. Because of perseverance of other pathways bypassing vSMC and controlling the initiation of nonverbal vocalizations (Fig. 1), such patients are occasionally able to initiate grunts, wails, and laughs, but they do not succeed in voluntary modulations of pitch, intensity, and the harmonious quality of their vocalizations (Simonyan and Horwitz, 2011). On the other hand, vSMC lesions in nonhuman primates have almost no effects on their vocalizations (Jürgens et al., 1982), which further suggests a highly specialized role of this region in the control of learned vocalizations, such as speech.

Our current conception of the sensorimotor cortex is heavily influenced by the homunculus model popularized by Wilder Penfield (Penfield and Boldrey, 1937). In the classic model, several key principles have defined our knowledge about the cortical representation of movement: (1) the precentral and postcentral gyri cleanly delineate motor from sensory functions, respectively; (2) an orderly topographic parcellation of brain regions corresponds to adjacent structures of the body; and (3) a particular body part or muscle maps one-to-one to the corresponding cortical site. The concept persisting for several decades regarding the vSMC organization featured a highly stereotyped, discretely ordered progression of representations for the lips, vocalization, jaw, tongue, and swallowing, respectively, along the dorsal-to-

ventral extent of the central sulcus (Fig. 2A). However, over the past several years, electrocortical stimulation mapping as well as neurophysiological recordings have revealed that such somatotopical organization may be an oversimplification, especially in the context of speaking.

A recent observational ECoG study characterized the individual variability across dozens of neurosurgical patients by providing a granular probabilistic description of evoked behavioral responses from stimulation of the vSMC (Breshears et al., 2015). This study found that mapping in a single individual rarely recapitulates Penfield's motor and sensory homunculi. Rather, some motor and sensory responses observed in one individual may be completely absent in another. Of note, these responses evoked by high-intensity stimulation were not those of voluntary natural movements or sensations. One possible explanation is that responses to suprathreshold currents reveal intrinsic "synergies" in muscle coordination. This is consistent with high-intensity microstimulation experiments in monkey SMC that resulted in complex, behaviorally relevant movements instead of single muscle group contractions (Graziano et al., 2002). Further support for this notion comes from the demonstration that microstimulation-evoked electromyographic patterns in macaques can be decomposed into smaller sets of muscle synergies that closely mirror those generated by natural hand movements (Overduin et al., 2012).

Breshears et al. (2015) further identified that cortical regions representing separate, but neighboring, body parts occupy overlapping regions of cortex such that a given point on vSMC may fall within the region for several, neighboring body parts (Fig. 2B). Generally, there is a bias for motor responses on the precentral gyrus and somatosensory responses on the postcentral gyrus as originally shown on Penfield's homunculi, but in practice both response types are found on both gyri. Some examples of motor responses evoked by cortical stimulation are contralateral pulling of the mouth, twitching of the lips, simple opening or closing of the mouth, or swallowing. Sensory responses are usually reported as tingling in a given body part, sometimes with extreme precision. These response types appear to be quite stereotyped across patients. Responses rarely, if ever, correspond to proprioceptive sensation or the perception of movement.

Although stimulation mapping has been foundational for understanding some of the basic organization of the vSMC, it is still unclear how these results extrapolate to the actual control of speech articulation. For example, unlike the unnatural and simple movements of single articulators evoked by electrical stimulation, the production of meaningful speech sounds requires the precisely coordinated control of multiple articulators, and thus meaningful speech production has not been evoked by focal electrical stimulation (Breshears et al., 2015). Instead, neurophysiological studies have leveraged the variability in articulatory patterns associated with the production of a large number of consonant-vowel syllables to quantitatively assign a dominant articulator (lips, jaw, tongue, or larynx) representation to the cortical activity recorded at each electrode (Bouchard et al., 2013, 2016). Although articulator representations appear to be partially overlapping in both space and time, a detailed dorsal-to-ventral organization of articulator representations has been identified (Fig. 2C,D). This is largely concordant with the results from stimulation mapping; however, two separate representations related to voicing from the larynx, with one site located ventral to the tongue and the other dorsal to the lips, were also identified.

Collectively, these results have revealed that vSMC is more complex than previously appreciated. The distinction between sensory and motor representations is blurred, and individ-

ual articulator representations appear to be interdigitated and overlapping. There is a general somatotopic mapping, but there is tremendous variability across individuals and fractured organization. Currently, it is unclear whether vSMC neural activity represents movement kinematics (Bouchard et al., 2016), acoustic targets of vocal production, or alternatively more complex features, such as movement trajectories or gestures. More research is needed to better define the nature of vSMC-driven movement representations and dynamics.

### The interplay between the vSMC and other cortical regions controlling speech production

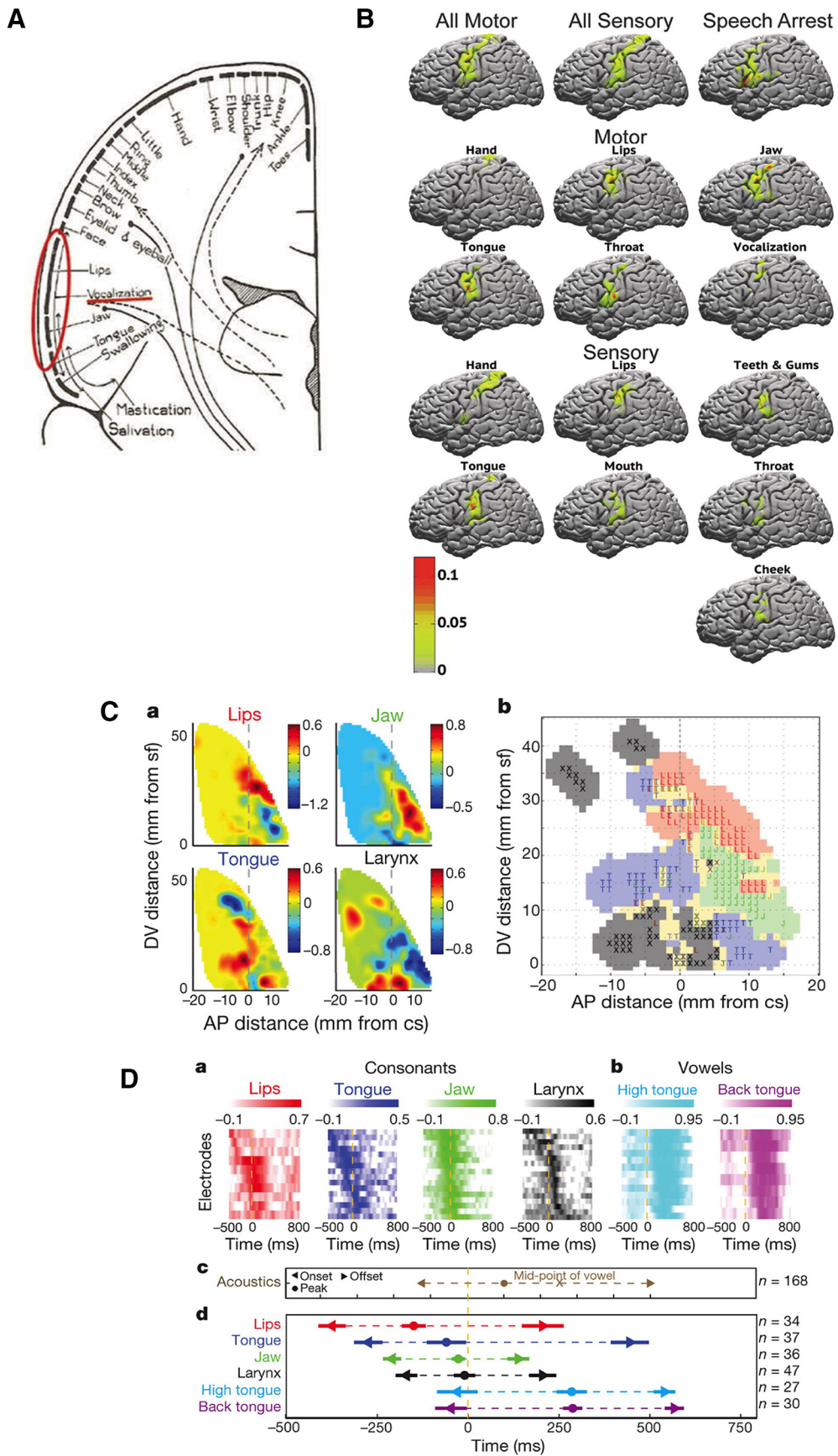
Recent technological advances in invasive human brain mapping introduced such experimental modalities as simultaneous field mapping, electrical stimulation tract tracing, and reversible cortical perturbation, which provided further insights not only into the organization of vSMC but also into the neural mechanisms underpinning the interplay between key brain regions involved in the control of speech production.

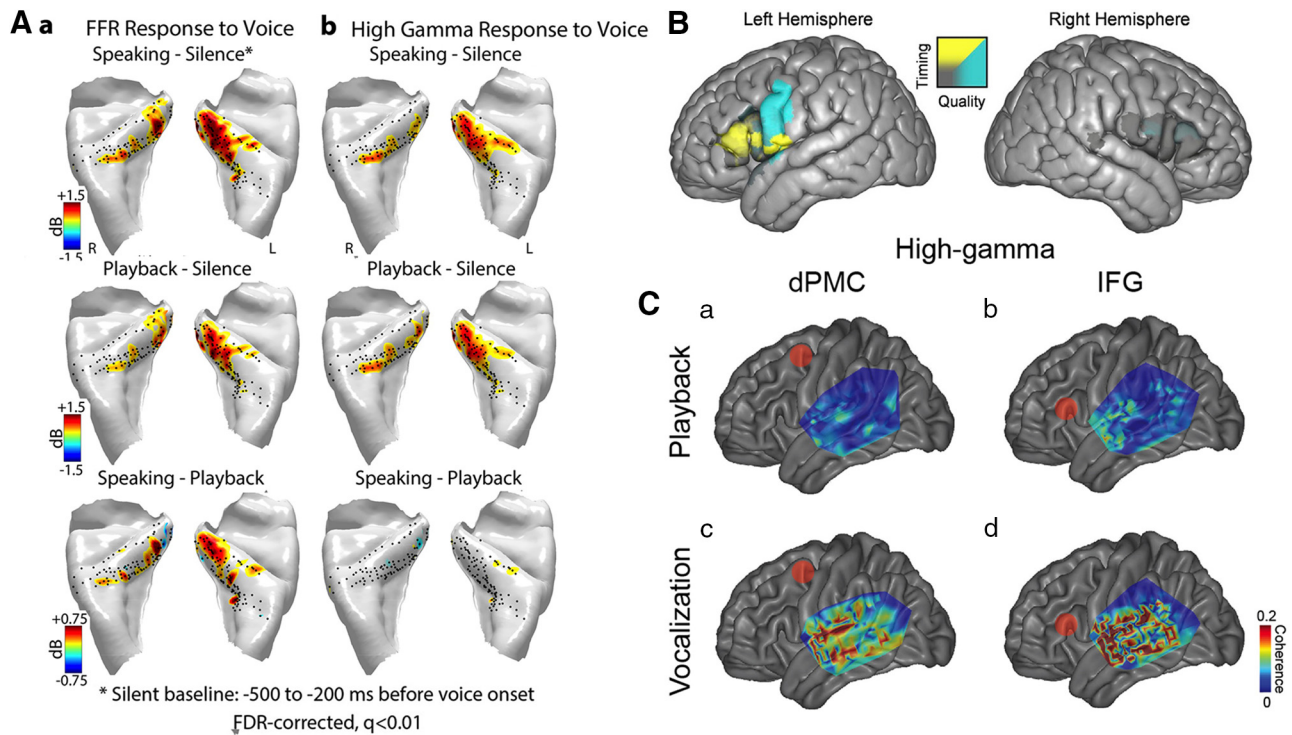
One of the important aspects of speaking is the audio-motor interactions and integration. "Auditory error cells" are hypothesized to reside in the posterior superior temporal gyrus (STG) (Guenther and Hickok, 2015), whereas human primary auditory cortex is located in the posteromedial Heschl's gyrus (HG). Leveraging the spatial extent of ECoG multielectrode arrays, a recent study has identified two distinct neural responses during speech production in HG: a frequency following response (FFR) and high-gamma (70–150 Hz) response to voice fundamental frequency (F0) (Behroozmand et al., 2016). The FFR was observed in both hemispheres and was modulated by speech production, with greater FFR amplitude during speaking compared with playback (Fig. 3Aa). Similar FFRs to voice F0 have not been seen in posterolateral STG (Flinker et al., 2010; Greenlee et al., 2011), suggesting different roles for vocal monitoring and error correction between primary and nonprimary auditory cortices. Conversely, high-gamma responses on HG to voice F0 did not show any modulation (Fig. 3Ab). Of note, the lack of high-gamma modulation on HG was also different from responses recorded from posterolateral STG (Flinker et al., 2010; Greenlee et al., 2011).

Another important regional contributor to the motor control of speech production is the inferior frontal gyrus (IFG). In a large cohort of neurosurgical patients, a recent study focally and reversibly perturbed brain function with brain surface cooling during awake craniotomy to detail the differential roles of the IFG (specifically, the Broca's area) and vSMC during speech production (Long et al., 2016). Perturbation of left IFG function resulted in alterations of speech timing, most commonly observed as speech slowing, whereas perturbation of the right IFG did not alter speech timing (Fig. 3B). Conversely, disruption of left vSMC function produced degradation in speech quality without changes in timing. Given the very focal nature of cortical perturbation in surface cooling, this study provided direct evidence for a specific role of Broca's area in the timing of speech sequences.

While these studies have elucidated the contribution of particular brain regions within the speech motor production network, recent series of ECoG-based electrical stimulation tract tracing studies have further revealed functional connections within speech motor regions. Functional coupling has been described between IFG and vSMC (Greenlee et al., 2004), within subregions of IFG (Greenlee et al., 2007), between primary and higher-order auditory areas on the posterolateral STG (Brugge et al., 2003), and between IFG and posterolateral STG (Garell et al.,







**Figure 3.** Differences in voice frequency following response (**Aa**) and high-gamma response (**Ab**) between primary auditory cortical areas on posteromedial HG compared with nonprimary areas on anterolateral HG. Data from Behroozmand et al. (2016). **B**, Changes in speech timing (yellow) versus quality (blue) resulting from focal brain cooling of the IFG and vSMC. Data from Long et al. (2016). **C**, Average coherence between auditory areas on lateral STG and dorsal premotor cortex (**a**, **c**) and IFG (**b**, **d**). Data from Kingyon et al. (2015).

2013). Although electrical stimulation tract tracing does not elucidate the anatomical connections between two functionally coupled areas, latency measurements of evoked responses can indicate the likely presence of a direct corticocortical component of the functional connection.

Another ECoG measure of functional connectivity is coherence, which can be computed based on simultaneously recorded time series from different brain sites (Swann et al., 2012). A recent study has identified coherence differences of posterior STG sites as a function of task (speaking vs playback), frequency band (theta vs high-gamma), and frontal brain region (dorsal premotor cortex vs IFG) (Kingyon et al., 2015). More specifically, during speech production, coherence was larger than that during playback, and coherence increased between STG and IFG compared with the coherence between STG and dorsal premotor cortex (Fig. 3C). Together, the presence of these functional connections outlines a mechanism for the postulated feedforward and feedback projections during speech production.

←

**Figure 2.** **A**, Schematic view of human body representation within the motor cortex (“motor homunculus”). Data from Penfield and Boldrey (1937). **B**, Probabilistic maps of the vSMC demonstrating the probability of observing a particular motor and sensory response as well as speech arrest to electrical stimulation at a particular cortical site. Color scale represents the probability of each response. Data from Breshears et al. (2015). **Ca**, Spatial localization of lips, jaw, tongue, and larynx representations within the vSMC. Average magnitude of articulator weightings (color scale) plotted as a function of anteroposterior (AP) distance from the central sulcus and dorsoventral (DV) distance from the Sylvian fissure. **Cb**, Functional somatotopic organization of speech-articulator representations in the vSMC. Red represents lips. Green represents jaw. Blue represents tongue. Black represents larynx. Yellow represents mixed. **D**, Timing of correlations between cortical activity and consonant (**Da**) and vowel (**Db**) articulator features with (**Dc**) acoustic landmarks, (**Dd**) temporal sequence, and range of correlations. Data from Bouchard et al. (2013).

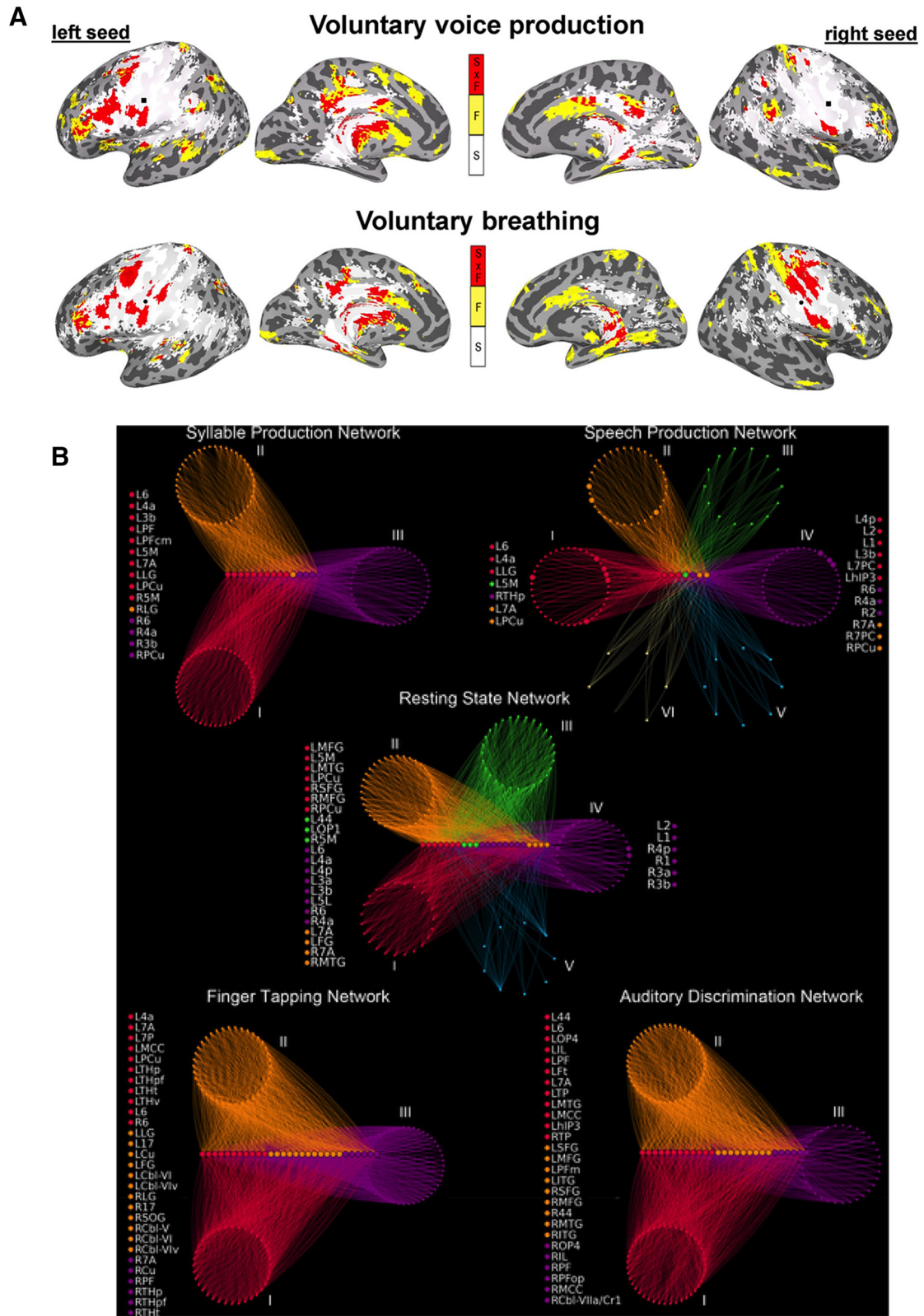
**Large-scale neural networks of speech production**

Although ECoG studies were successful in advancing our knowledge about the detailed organization of the vSMC for speech motor control and the interplay between specific brain regions within the speech production network, a recent series of fMRI and diffusion-weighted tractography studies were instrumental in identifying the large-scale neural network architecture of speech sensorimotor control.

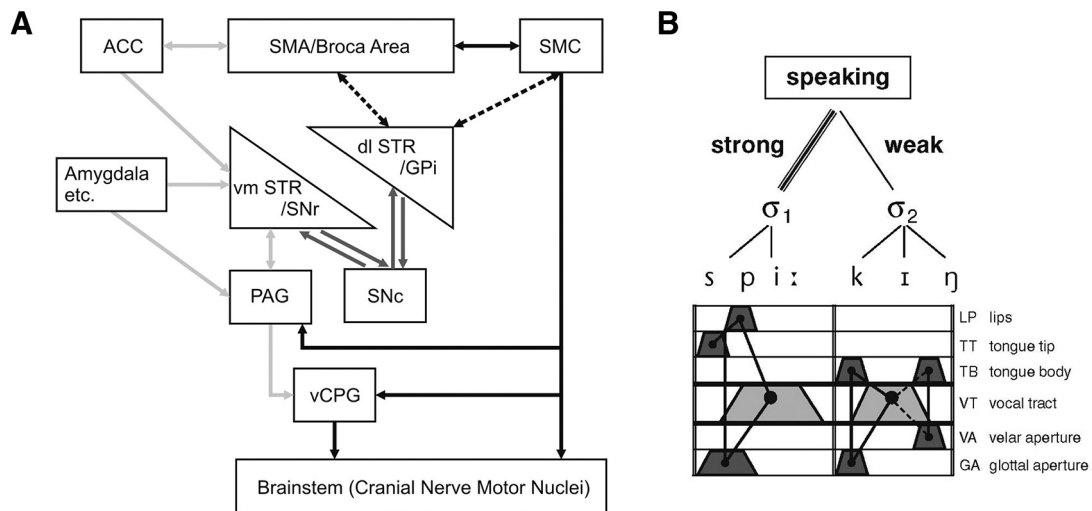
Although these noninvasive studies lack the temporal and spatial resolution of ECoG studies, they largely agreed on the localization of different components of speech articulator representations within the vSMC (for meta-analyses, see Takai et al., 2010; Simonyan, 2014). Follow-up tractography studies have further identified a predominantly bilateral structural network originating from the speech motor cortex, upon which different functional networks are built to control various components of speech motor output, such as syllable production and voluntary breathing (Simonyan et al., 2009; Kumar et al., 2016) (Fig. 4A). Importantly, the laryngeal motor cortex was found to establish nearly sevenfold stronger structural connectivity with the somatosensory and inferior parietal cortices in humans compared with nonhuman primates (Kumar et al., 2016). In agreement with ECoG functional connectivity studies described above, these findings suggest that the evolution of enhanced motocortical-parietal connections likely allowed for more complex sensorimotor coordination and modulation of learned vocalizations for speech production.

In addition to analyzing the role of a particular brain region and its specific long-range connections within the speech controlling network, it is important to consider that a spoken word requires the orchestration of multiple neural networks associated with various speech-related processes, including sound percep-





**Figure 4.** *A*, Common and distinct functional and structural networks of the laryngeal motor cortex during syllable production and voluntary breathing. Yellow represents functional connections (F) underlying each task. White represents structural connections (S) underlying each task. Red represents overlap between the functional and structural connections (FoS). Data from Simonyan et al. (2009). *B*, Functional community structure of the group-averaged networks during the resting state, syllable production, sentence production, sequential finger tapping, and auditory discrimination of pure tones. Distinct network communities are shown as circular groups of nodes positioned around the respective connector hubs, which are arranged on horizontal lines. Nodal colors represent module membership. Node lists on the left and right of each graph indicate connector and provincial hubs, respectively. 1, area 1; 17, area 17; 2, area 2; 3a/3b, areas 3a/3b; 44, area 44; 4a/4p, anterior/posterior part of area 4; 5L/5M, area 5L/5M; 6, area 6; 7A/7P/7PC, area 7A/7P/7PC; Cbl-VI/VIv/VIIa/Cr1, cerebellar lobules VI/VIv/VIIa/Cr1; Cu, cuneus; FG, fusiform gyrus; hIP3, areas hIP3; IL, insula; SOG, superior occipital gyrus; ITG/MTG, inferior/middle temporal gyrus; LG, lingual gyrus; MCC, middle cingulate cortex; OP1–4, operculum; PCu, precuneus; PF/PFm/PFop/PFp/PGa/PgP, areas PF/PFm/PFop/PFp/PGa/PgP in the inferior parietal cortex; MFG, middle frontal gyrus; Thp/Thpf/Thpm/ThT, parietal/prefrontal/premotor/temporal part of the thalamus; TP, temporal pole; R, right; L, left. Data from Fuertinger et al. (2015).



**Figure 5.** *A*, The cerebral networks supporting primate-general (gray arrows) and human-specific (black) aspects of vocal communication are assumed to be closely intertwined at the level of the basal ganglia. Dashed lines indicate that the basal ganglia motor loop undergoes a dynamic ontogenetic reorganization during spoken language acquisition in that a left-hemisphere cortical storage site of syllable-sized motor programs gradually emerges. Amygdala etc., Amygdala and other structures of the limbic system; ACC, anterior cingulate cortex; SMA, supplementary motor area; GPI, internal segment of globus pallidus; SNr/SNc, substantia nigra, pars reticulata/pars compacta; PAG, periaqueductal gray; vCPG, vocal central pattern generator. Data from Ackermann et al. (2014). *B*, Gestural architecture of the word “speaking.” Laryngeal activity (bottom line) is a crucial part of the respective movement sequence and must be adjusted to other vocal tract excursions. Articulatory gestures are assorted into syllabic units; gesture bundles pertaining to strong and weak syllables are rhythmically patterned to form metrical feet. Data from Ziegler (2010).

tion, semantic processing, memory encoding, preparation, and execution of vocal motor commands (e.g., Hickok and Poeppel, 2007; Houde and Nagarajan, 2011; Tourville and Guenther, 2011; Price, 2012). However, a number of questions about how and where these large-scale brain networks interact with one another remained open until recently. Using inter-regional functional connectivity analysis from seven key brain regions controlling speech (i.e., vSMC, IFG, STG, supplementary motor area, cingulate cortex, putamen, and thalamus), a recent study has determined that the strongest interaction between individual networks during speech production is centered around the bilateral vSMC, IFG, and supplementary motor area as well as the right STG (Simonyan and Fuertinger, 2015). Among the examined networks, the vSMC (specifically, its laryngeal region) establishes a common core network that fully overlaps with all other speech-related networks, determining the extent of network interactions. On the other hand, the inferior parietal lobule and cerebellum are the most heterogeneous regions preferentially recruited into the functional speech network and facilitating the transition from the resting state to speaking.

The complexity of the speech production network was further examined using a multivariate graph theoretical analysis of fMRI data in healthy humans by constructing functional networks of increasing hierarchy from the resting state to the motor output of meaningless syllables to the production of complex real-life speech as well as compared with non-speech-related finger tapping and pure tone discrimination networks (Fuertinger et al., 2015). This study demonstrated the intricate involvement of the vSMC in the control of speech production. Specifically, the presence of a segregated network of highly connected local neural communities of information transfer (i.e., hubs) was found in the vSMC and inferior parietal lobule, which formed a shared core hub network that was common to all examined conditions (Fig. 4*B*). Importantly, this SMC-centered core network exhibited features of multimodal flexible hubs similar to those found in frontoparietal brain regions (Cole et al., 2013) by adaptively switching its long-range functional connectivity depending on the task content, which resulted in the formation of distinct neural commu-

nities characteristic for each task (Fuertinger et al., 2015) (Fig. 4*B*). The speech production network exhibited the emergence of the left primary motor cortex as a particularly influential hub as well as by the full integration of the prefrontal cortex, insula, putamen, and thalamus, which were less important for other examined networks, including the closely related syllable production network. Collectively, the specialized rearrangement of the global network architecture shaped the formation of the functional speech connectome, whereas the capacity of the SMC for operational heterogeneity challenged the long-established concept of low-order unimodality of this region.

### The contribution of the basal ganglia and cerebellum to the control of speech production

In addition to the vSMC and other cortical areas, speech production recruits the basal ganglia and cerebellum. Among the subcortical structures, the laryngeal/orofacial motor cortex establishes the strongest direct connections with different functional subdivisions of the basal ganglia that are engaged in sensorimotor control of movement coordination and execution (posterior dorsal striatum), cognitive processing (anterior dorsal putamen), and attention and memory processing (basal nucleus of Meynert) (Künzle, 1975; Jürgens, 1976; Simonyan and Jürgens, 2003, 2005). Striatal lesions are known to cause speech motor disturbances, including dysphonia, dysarthria, and other verbal aphasia, while these lesions have no profound effects on monkey vocalizations (e.g., Damasio et al., 1982; Jürgens et al., 1982; Nadeau and Crosson, 1997). This suggests that the striatum may specifically be involved in the control of learned voice production (Jürgens, 2002; Simonyan et al., 2012; Ackermann et al., 2014). Furthermore, as evident from clinical data in neurological patients, bilateral prenatal and perinatal damage to the striatum leads to compromised acquisition of fluent speech utterances, which is in stark contrast with speech motor deficits (e.g., monotone, hypotonic speech, reduced loudness and pitch, decreased articulatory accuracy) that are seen in adult patients with Parkinson’s disease or cerebrovascular disorders, such as stroke (Ackermann et al., 2014). It is therefore plausible to suggest that the striatum might be critical for the initial organization of speech motor programs. Speech deficits



observed in patients with movement disorders and cerebrovascular diseases further indicate that damage to the basal ganglia leads not only to disturbances of motor control but also to emotive-prosodic modulation of the sound structure of verbal utterances (Ackermann et al., 2014). The capability of a single one-dimensional speech wave to simultaneously convey both the propositional and emotional contents of spoken language is thought to be based on dopamine-dependent cascading interconnections between limbic and motor basal ganglia loops (Haber, 2010) as well as on the convergence of the descending voice controlling motor and limbic pathways at the level of the basal ganglia (Fig. 5A).

Another important contribution of basal ganglia control is the neurochemical modulation of speech production. Endogenous dopamine release in the left ventromedial portion of the associative striatum has been shown to be coupled with neural activity during speaking, influencing the left-hemispheric lateralization of the functional speech network (Simonyan et al., 2013). Greater involvement of the goal-directed associative striatum suggests that dopaminergic influences on cognitive aspects of speech control weigh in more significantly for information processing during ongoing speaking. On the other hand, modulatory effects of dopaminergic function in the habitual sensorimotor striatum may prevail in the course of speech and language development as well as during the acquisition of a second language, which requires higher integration of the sensorimotor system for shaping novel articulatory sequences (Simonyan et al., 2013).

In contrast to the basal ganglia, the cerebellum generally engages in movement preparation and execution as well as motor skill acquisition, including those for speech production, although the underlying mechanisms remain to be further elucidated. Cerebellar disorders may give rise to the syndrome of ataxic dysarthria that are characterized by compromised stability of sound production and slowed execution of single articulatory gestures, especially under enhanced temporal constraints (Ackermann and Brendel, 2016). These abnormalities accord well with the pathophysiological deficits observed in upper limb ataxia. Most noteworthy, reduced maximum speaking rate appears to approach a plateau at ~2.5–3 Hz in patients with a purely cerebellar disorder. Therefore, the processing capabilities of the cerebellum seem to provide a necessary prerequisite to push speaking rate beyond this level and, thus, to engage in the modulation of the rhythmic structure of verbal utterances. For example, the length of successive syllables has to be adjusted to metrical and rhythmic demands (Fig. 5B). Because, from a phylogenetic perspective, inner speech mechanisms (i.e., prearticulatory verbal code) may have emerged from overt speech, the computational power of the cerebellum might also enable certain aspects of the sequential organization of prearticulatory verbal codes. Cerebellar disorders, therefore, may compromise cognitive operations associated with “inner speech,” such as the linguistic scaffolding of executive functions.

In conclusion, research in the past few years has come a long way in localizing, mapping, and providing mechanistic explanations of some of the fundamental principles of speech motor cortical organization. With studies leveraging methodological advances and developing novel modalities in brain mapping, it is an exciting time in the field of speech production research, which will continue to challenge empirical concepts and potentially outline new directions for elucidation of long-envisioned neural mechanisms of human speech. Investigations of the neural bases of speech production are important not only for understanding the basic principles of speaking but also have a high clinical relevance. Speech-related disability is frequently associated with major neurological and psychiatric problems, such as Parkinson’s

disease, stuttering, spasmodic dysphonia, stroke, and schizophrenia, to name a few. Thus, a lack of knowledge about normal speech motor control may have a long-term impact on our ability to understand speech motor disturbances in these disorders. Therefore, the continuous investigation of brain mechanisms underlying normal speech production is critically important for the development of new translational approaches to address the unanswered questions about speech alterations in a wide range of human brain disorders.

## References

- Ackermann H, Brendel B (2016) Cerebellar contributions to speech and language. In: *Neurobiology of language* (Hickok G, Small SL, eds), pp 73–84. Amsterdam: Elsevier/Academic.
- Ackermann H, Hage SR, Ziegler W (2014) Brain mechanisms of acoustic communication in humans and nonhuman primates: an evolutionary perspective. *Behav Brain Sci* 37:529–546. [CrossRef Medline](#)
- Arbib MA, Liebal K, Pika S (2008) Primate vocalization, gesture, and the evolution of human language. *Curr Anthropol* 49:1053–1063; discussion 1063–1076. [Medline](#)
- Behroozmand R, Oya H, Nourski KV, Kawasaki H, Larson CR, Brugge JF, Howard MA 3rd, Greenlee JD (2016) Neural correlates of vocal production and motor control in human Heschl’s gyrus. *J Neurosci* 36:2302–2315. [CrossRef Medline](#)
- Bouchard KE, Mesgarani N, Johnson K, Chang EF (2013) Functional organization of human sensorimotor cortex for speech articulation. *Nature* 495:327–332. [CrossRef Medline](#)
- Bouchard KE, Conant DF, Anumanchipalli GK, Dichter B, Chaisanguanthum KS, Johnson K, Chang EF (2016) High-resolution, non-invasive imaging of upper vocal tract articulators compatible with human brain recordings. *PLoS One* 11:e0151327. [CrossRef Medline](#)
- Breshears JD, Molinaro AM, Chang EF (2015) A probabilistic map of the human ventral sensorimotor cortex using electrical stimulation. *J Neurosurg* 123:340–349. [CrossRef Medline](#)
- Brugge JF, Volkov IO, Garell PC, Reale RA, Howard MA 3rd (2003) Functional connections between auditory cortex on Heschl’s gyrus and on the lateral superior temporal gyrus in humans. *J Neurophysiol* 90:3750–3763. [CrossRef Medline](#)
- Callan DE, Tsytarev V, Hanakawa T, Callan AM, Katsuhara M, Fukuyama H, Turner R (2006) Song and speech: brain regions involved with perception and covert production. *Neuroimage* 31:1327–1342. [CrossRef Medline](#)
- Cole MW, Reynolds JR, Power JD, Repovs G, Anticevic A, Braver TS (2013) Multi-task connectivity reveals flexible hubs for adaptive task control. *Nat Neurosci* 16:1348–1355. [CrossRef Medline](#)
- Damasio AR, Damasio H, Rizzo M, Varney N, Gersh F (1982) Aphasia with nonhemorrhagic lesions in the basal ganglia and internal capsule. *Arch Neurol* 39:15–24. [CrossRef Medline](#)
- Flinker A, Chang EF, Kirsch HE, Barbaro NM, Crone NE, Knight RT (2010) Single-trial speech suppression of auditory cortex activity in humans. *J Neurosci* 30:16643–16650. [CrossRef Medline](#)
- Fuertinger S, Horwitz B, Simonyan K (2015) The functional connectome of speech control. *PLoS Biol* 13:e1002209. [CrossRef Medline](#)
- Garell PC, Bakken H, Greenlee JD, Volkov I, Reale RA, Oya H, Kawasaki H, Howard MA, Brugge JF (2013) Functional connection between posterior superior temporal gyrus and ventrolateral prefrontal cortex in human. *Cereb Cortex* 23:2309–2321. [CrossRef Medline](#)
- Gemba J (2002) Motor programming for hand and vocalization movements. In: *Principles of frontal lobe function* (Stuss DT, Knight RT, eds), pp 127–148. New York: Oxford UP.
- Graziano MS, Taylor CS, Moore T (2002) Complex movements evoked by microstimulation of precentral cortex. *Neuron* 34:841–851. [CrossRef Medline](#)
- Greenlee JD, Oya H, Kawasaki H, Volkov IO, Kaufman OP, Kovach C, Howard MA, Brugge JF (2004) A functional connection between inferior frontal gyrus and orofacial motor cortex in human. *J Neurophysiol* 92:1153–1164. [CrossRef Medline](#)
- Greenlee JD, Oya H, Kawasaki H, Volkov IO, Severson MA 3rd, Howard MA 3rd, Brugge JF (2007) Functional connections within the human inferior frontal gyrus. *J Comp Neurol* 503:550–559. [CrossRef Medline](#)
- Greenlee JD, Jackson AW, Chen F, Larson CR, Oya H, Kawasaki H, Chen H,

- Howard MA 3rd (2011) Human auditory cortical activation during self-vocalization. *PLoS One* 6:e14744. [CrossRef Medline](#)
- Guenther FH, Hickok G (2015) Role of the auditory system in speech production. *Handb Clin Neurol* 129:161–175. [CrossRef Medline](#)
- Haber SN (2010) Convergence of limbic, cognitive, and motor corticostriatal circuits with dopamine pathways in primate brain. In: *Dopamine handbook* (Iversen LL, Dunnett SB, Björklund A, ed), pp 38–48. Oxford: Oxford UP.
- Hage SR (2010) Neuronal networks involved in the generation of vocalization. In: *Handbook of mammalian vocalization: an integrative neuroscience approach* (Brudzynski SM, ed), pp 339–350. Amsterdam: Elsevier.
- Hage SR, Gavrilov N, Salomon F, Stein AM (2013) Temporal vocal features suggest different call-pattern generating mechanisms in mice and bats. *BMC Neurosci* 14:99. [CrossRef Medline](#)
- Hickok G, Poeppel D (2007) The cortical organization of speech processing. *Nat Rev Neurosci* 8:393–402. [CrossRef Medline](#)
- Hickok G, Houde J, Rong F (2011) Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron* 69:407–422. [CrossRef Medline](#)
- Houde JF, Nagarajan SS (2011) Speech production as state feedback control. *Front Hum Neurosci* 5:82. [CrossRef Medline](#)
- Jürgens U (1976) Projections from the cortical larynx area in the squirrel monkey. *Exp Brain Res* 25:401–411. [Medline](#)
- Jürgens U (2002) Neural pathways underlying vocal control. *Neurosci Biobehav Rev* 26:235–258. [CrossRef Medline](#)
- 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. [CrossRef Medline](#)
- Kawai R, Markman T, Poddar R, Ko R, Fantana AL, Dhawale AK, Kampff AR, Ölveczky BP (2015) Motor cortex is required for learning but not for executing a motor skill. *Neuron* 86:800–812. [CrossRef Medline](#)
- Kingyon J, Behroozmand R, Kelley R, Oya H, Kawasaki H, Narayanan NS, Greenlee JD (2015) High-gamma band fronto-temporal coherence as a measure of functional connectivity in speech motor control. *Neuroscience* 305:15–25. [CrossRef Medline](#)
- Kirzinger A, Jürgens U (1982) Cortical lesion effects and vocalization in the squirrel monkey. *Brain Res* 233:299–315. [CrossRef Medline](#)
- Kumar V, Crosson PL, Simonyan K (2016) Structural organization of the laryngeal motor cortical network and its implication for evolution of speech production. *J Neurosci* 36:4170–4181. [CrossRef Medline](#)
- Künzle H (1975) Bilateral projections from precentral motor cortex to the putamen and other parts of the basal ganglia: an autoradiographic study in *Macaca fascicularis*. *Brain Res* 88:195–209. [CrossRef Medline](#)
- Lachaux JP, Axmacher N, Mormann F, Halgren E, Crone NE (2012) High-frequency neural activity and human cognition: past, present and possible future of intracranial EEG research. *Prog Neurobiol* 98:279–301. [CrossRef Medline](#)
- Lameira AR, Hardus ME, Mielke A, Wich SA, Shumaker RW (2016) Vocal fold control beyond the species-specific repertoire in an orangutan. *Sci Rep* 6:30315. [CrossRef Medline](#)
- Long MA, Katlowitz KA, Svirsky MA, Clary RC, Byun TM, Majaj N, Oya H, Howard MA 3rd, Greenlee JD (2016) Functional segregation of cortical regions underlying speech timing and articulation. *Neuron* 89:1187–1193. [CrossRef Medline](#)
- Nadeau SE, Crosson B (1997) Subcortical aphasia. *Brain Lang* 58:355–402; discussion 418–423. [CrossRef Medline](#)
- Overduin SA, d'Avella A, Carmena JM, Bizzi E (2012) Microstimulation activates a handful of muscle synergies. *Neuron* 76:1071–1077.
- Penfield W, Boldrey E (1937) Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain* 60:389–443. [CrossRef](#)
- Poeppel D, Emmorey K, Hickok G, Pylkkänen L (2012) Towards a new neurobiology of language. *J Neurosci* 32:14125–14131. [CrossRef Medline](#)
- Price CJ (2012) A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage* 62:816–847. [CrossRef Medline](#)
- Simonyan K (2014) The laryngeal motor cortex: its organization and connectivity. *Curr Opin Neurobiol* 28:15–21. [CrossRef Medline](#)
- Simonyan K, Fuertinger S (2015) Speech networks at rest and in action: interactions between functional brain networks controlling speech production. *J Neurophysiol* 113:2967–2978. [CrossRef Medline](#)
- Simonyan K, Horwitz B (2011) Laryngeal motor cortex and control of speech in humans. *Neuroscientist* 17:197–208. [CrossRef Medline](#)
- Simonyan K, Jürgens U (2003) Efferent subcortical projections of the laryngeal motor cortex in the rhesus monkey. *Brain Res* 974:43–59. [CrossRef Medline](#)
- Simonyan K, Jürgens U (2005) Afferent subcortical connections into the motor cortical larynx area in the rhesus monkey. *Neuroscience* 130:119–131. [CrossRef Medline](#)
- Simonyan K, Ostuni J, Ludlow CL, Horwitz B (2009) Functional but not structural networks of the human laryngeal motor cortex show left hemispheric lateralization during syllable but not breathing production. *J Neurosci* 29:14912–14923. [CrossRef Medline](#)
- Simonyan K, Horwitz B, Jarvis ED (2012) Dopamine regulation of human speech and bird song: a critical review. *Brain Lang* 122:142–150. [CrossRef Medline](#)
- Simonyan K, Herscovitch P, Horwitz B (2013) Speech-induced striatal dopamine release is left lateralized and coupled to functional striatal circuits in healthy humans: a combined PET, fMRI and DTI study. *Neuroimage* 70:21–32. [CrossRef Medline](#)
- Swann NC, Cai W, Conner CR, Pieters TA, Claffey MP, George JS, Aron AR, Tandon N (2012) Roles for the pre-supplementary motor area and the right inferior frontal gyrus in stopping action: electrophysiological responses and functional and structural connectivity. *Neuroimage* 59:2860–2870. [CrossRef Medline](#)
- Takahashi DY, Fenley AR, Teramoto Y, Narayanan DZ, Borjon JI, Holmes P, Ghazanfar AA (2015) LANGUAGE DEVELOPMENT: the developmental dynamics of marmoset monkey vocal production. *Science* 349:734–738. [CrossRef Medline](#)
- Takai O, Brown S, Liotti M (2010) Representation of the speech effectors in the human motor cortex: somatotopy or overlap? *Brain Lang* 113:39–44. [CrossRef Medline](#)
- Tankus A, Fried I, Shoham S (2014) Cognitive-motor brain-machine interfaces. *J Physiol Paris* 108:38–44. [CrossRef Medline](#)
- Tourville JA, Guenther FH (2011) The DIVA model: a neural theory of speech acquisition and production. *Lang Cogn Process* 26:952–981. [CrossRef Medline](#)
- Winter P, Handley P, Ploog D, Schott D (1973) Ontogeny of squirrel monkey calls under normal conditions and under acoustic isolation. *Behaviour* 47:230–239. [CrossRef Medline](#)
- Ziegler W (2010) Apraxic failure and the hierarchical structure of speech motor plans: A nonlinear probabilistic model. In: *Assessment of Motor Speech Disorders* (Lowit A, Kent RD, eds), pp 305–323. San Diego: Plural Publishing.