



Beyond Broca: neural architecture and evolution of a dual motor speech coordination system

Gregory Hickok,^{1,2} Jonathan Venezia^{3,4} and Alex Teghipco⁵

Classical neural architecture models of speech production propose a single system centred on Broca's area coordinating all the vocal articulators from lips to larynx. Modern evidence has challenged both the idea that Broca's area is involved in motor speech coordination and that there is only one coordination network. Drawing on a wide range of evidence, here we propose a dual speech coordination model in which laryngeal control of pitch-related aspects of prosody and song are coordinated by a hierarchically organized dorsolateral system while supralaryngeal articulation at the phonetic/syllabic level is coordinated by a more ventral system posterior to Broca's area. We argue further that these two speech production subsystems have distinguishable evolutionary histories and discuss the implications for models of language evolution.

- 1 Department of Cognitive Sciences, University of California, Irvine, CA 92697, USA
- 2 Department of Language Science, University of California, Irvine, CA 92697, USA
- 3 Auditory Research Laboratory, VA Loma Linda Healthcare System, Loma Linda, CA 92357, USA
- 4 Department of Otolaryngology—Head and Neck Surgery, Loma Linda University School of Medicine, Loma Linda, CA 92350, USA
- 5 Department of Psychology, University of South Carolina, Columbia, SC 29208, USA

Correspondence to: Gregory Hickok
Department of Language Science
University of California, Irvine, CA 92697, USA
E-mail: greg.hickok@uci.edu

Keywords: speech production; motor speech; auditory; sensorimotor control; precentral gyrus; evolution

Introduction

In a series of reports in the 1860s, Paul Broca famously implicated a single area in the posterior portion of the left inferior frontal gyrus—now called Broca's area—as the seat of the 'faculty of articulated language'.^{1–3} Broca was clear about what he meant by 'articulated language': 'it is not the faculty of language, it is not the memory for words, nor is it the action of the nerves and muscles corresponding to phonation and to articulation', rather, it is 'the faculty that coordinates the proper movement of articulated language'¹ (p. 332) (emphasis original; translation from Green⁴). Just over a century later, Norman Geschwind endorsed Broca's claims: 'Broca's area, which is adjacent to the region of the motor cortex that controls the movement of the muscles of the lips, the jaw, the tongue, the soft palate and the vocal cords, apparently

incorporates programs for the coordination of these muscles in speech'.⁵ (p. 78)

Modern evidence, however, has challenged the idea that Broca's area is involved in motor speech coordination, which we define here as a level of planning intermediate to lower-level articulatory control in primary motor cortex and higher-level morphosyntactic processes, and it has challenged the idea that there is only one coordination network. Regarding Broca's area, typically defined as the pars opercularis and pars triangularis of the inferior frontal gyrus, we now appreciate that it is not functionally or anatomically monolithic^{6–9} and that this heterogeneity has been linked not to speech coordination but to various higher-level processes including syntax,^{10–13} lexical selection,^{9,14} working memory^{8,15} and sequencing.^{16–18} In contrast, speech coordination deficits of the sort

Broca described, such as those found in apraxia of speech,¹⁹ have been associated primarily with more posterior frontal motor-related areas and the anterior insula.^{20–26} Regarding the idea of a single speech coordination network, many studies have reported not only ventral areas involved in speech tasks, but also a more dorsal speech region, which is consistently activated in simple receptive and expressive speech tasks^{17,27–29} and when directly stimulated can induce speech arrest³⁰ and vocalization,³¹ making this more dorsal zone an additional candidate for a speech coordination area. The possibility of a dual speech coordination architecture is the topic of this article.

In the following sections we motivate a new neural architecture model, which parcellates speech coordination into two parallel systems, one for pitch-related vocalization (prosody, song) predominantly involving laryngeal control, and the other for phonetic/syllabic articulation involving the supralaryngeal vocal tract. Our discussion is divided into two main parts. Part 1 reviews the evidence for two separable speech areas on the precentral gyrus, which motivates our hypothesized functional distinction. Part 2 considers *why* the human brain might have evolved two separable speech coordination systems.

Part 1: two speech coordination areas and a hypothesis

Two speech areas in the lateral precentral gyrus

Evidence for a more dorsal speech coordination area emerged as early as 1861 when Broca presented the neuroanatomical findings from his second case, Lelong, of speech dyscoordination or ‘aphemia’ as he called it.³² Lelong’s lesion was more focal than Broca’s first case as it involved only the most posterior part of the pars opercularis and extended dorsally to include the posterior middle frontal gyrus.³³ Broca took note of this fact: ‘The integrity of the third frontal convolution (*and perhaps of the second*) seems indispensable to the exercise of the faculty of articulate language’ (emphasis added, translation from Dronkers et al.³³). Modern-day evidence from a variety of sources provides strong support for the existence of two speech coordination areas in the precentral gyrus, which we summarize next.

Functional imaging

Functional imaging evidence began emerging in the 1990s for two speech areas in the posterior–lateral frontal cortex, a ventral site corresponding to the pars opercularis portion of Broca’s area and a more dorsal site in the precentral gyrus—likely premotor cortex on the gyral crown rather than primary motor cortex in the central sulcus—just posterior to the middle frontal gyrus^{27–29} (Fig. 1). This evidence was consistent enough to feature the two zones in modern dual-stream models of speech processing,^{34–36} although the functional roles of the two regions were and remain unclear.

Neurosurgical stimulation mapping

Further evidence for two lateral precentral gyrus speech coordination zones comes from a recent multisite retrospective study of the anatomical distribution of speech disruption during direct cortical stimulation mapping.³⁰ Disruption was classified as either (i) ‘speech arrest’, defined by ‘complete interruption while counting, without obvious oral, facial, mandibular, and laryngeal muscle movement’, which corresponds most closely to Broca’s aphemia;

or (ii) ‘anomia’, defined as the ‘inability to name the object in a picture, or misnaming by using the wrong word (e.g. using ‘cat’ instead of ‘dog’), while still being able to speak the leading word [e.g. ‘This is a(n) . . .’].’³⁰ (p. 5) Examining data from nearly 600 patients, four clusters associated with speech arrest were identified. The two most dense clusters were ventral posterior frontal cortex (Fig. 2A, green cluster #1) and dorsolateral posterior frontal cortex (yellow cluster #2), each of which had centroids in the precentral gyrus. The other two clusters involved the supplementary motor area and the posterior superior temporal lobe. Classical Broca’s area, including the pars opercularis, was not consistently implicated in speech arrest, but was implicated in anomia, consistent with other evidence suggesting a higher-level linguistic function for Broca’s area.

Apraxia of speech

Apraxia of speech (AOS) is a motor speech disorder that is argued to reflect an impairment in the planning and/or programming of speech, situated on a functional hierarchy in between higher-level aphasias and lower-level dysarthrias.^{19,39} As such, AOS is the modern-day disorder that most closely fits Broca’s (and our) definition of a speech coordination deficit. There are persistent debates about the underlying neuroanatomy of AOS, but it is noteworthy in the present context that some investigators emphasize ventral regions and the anterior insula,^{20,21} while others have found evidence for a more dorsal premotor focus^{22–24} or both,^{25,26} thus mirroring the dual-zone neurosurgical and functional imaging observations. There are further debates about whether there are subtypes of apraxia of speech.⁴⁰ One proposal that is gaining support is a distinction between phonetic (predominantly phoneme distortions, substitutions, additions) versus prosodic (predominantly suprasegmental distortions of rate, segmentation, prosody) variants.⁴⁰ Although much remains to be worked out in terms of the behavioural taxonomy and underlying neuroanatomy, perhaps different subtypes of AOS correlate with relative ventral versus dorsal brain damage. Indeed, a recent case report indicates that surgical resection of the posterior middle frontal gyrus can result in chronic pure apraxia of speech with prosodic dysfunction.²²

Speech repetition

Speech repetition has long been used as a diagnostic test for the integrity of auditory–motor speech circuits as it reduces the demand on lexical access and selection, while placing higher loads on more direct sensorimotor interaction.⁴¹ In the classical model, the neuroanatomy underlying this ability involves the arcuate fasciculus, a major white matter pathway thought to connect Wernicke’s area with Broca’s area.⁴² However, the picture is more complicated^{43,44} with connectivity from various posterior temporal and parietal regions to both dorsal and ventral speech zones.⁴⁵ A recent data-driven study directly assessed the region-to-region connectivity underlying speech repetition deficits in post-stroke aphasia, reporting that repetition deficits were predicted by disruption of white matter connectivity to both the ventral and dorsal frontal speech areas.³⁸ Neither of these connectivity patterns implicated an arcuate fasciculus-mediated connection between Wernicke’s and Broca’s areas. Instead, the critical connection to the ventral speech region was with inferior parietal secondary somatosensory cortex in the anterior supramarginal gyrus, while the critical connections to the dorsal speech area were with auditory cortical areas in the superior temporal gyrus (Fig. 2B). We discuss this difference in sensory connectivity in subsequent sections.

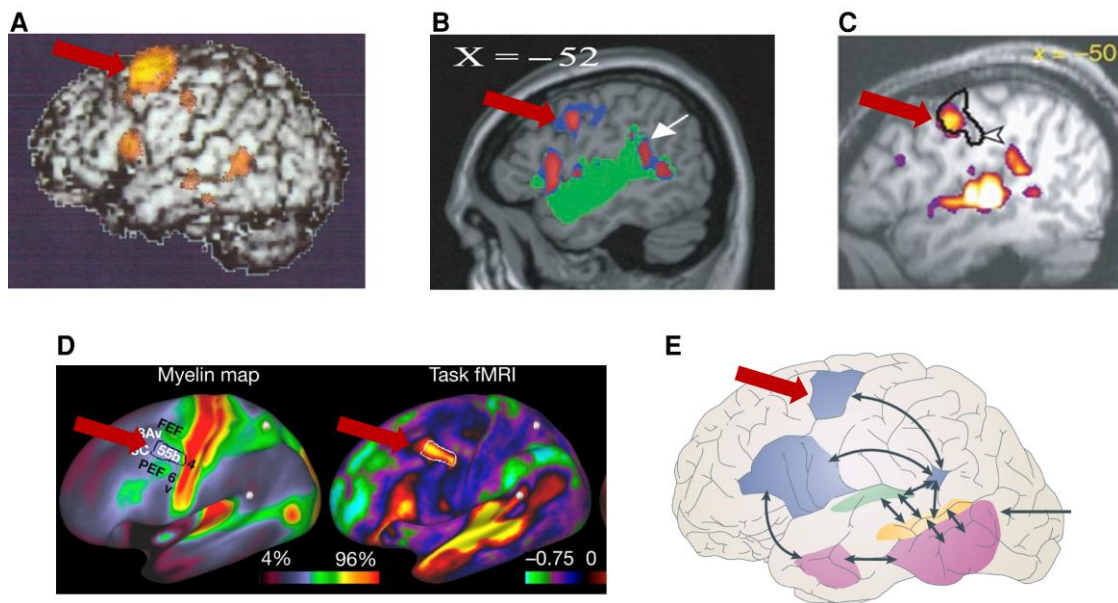


Figure 1 Brain maps from a range of studies implicating a dorsal precentral gyrus region in speech processing (red arrows). (A) PET study showing activation associated with repeating words minus saying ‘crime’ after hearing a reversed word.²⁷ (B) Functional MRI study showing activation associated with both auditory (listening to nonsense sentences) and motor (silently rehearsing nonsense sentences) speech.²⁹ (C) Functional MRI study showing activation associated with listening to syllables (heat map) and producing the same syllables (outline).²⁸ (D) Myelin (left) and functional MRI task activation (right) map from the Human Connectome Project database showing area 55b.³⁷ (E) Schematic diagram of the dual stream model of speech processing showing two frontal regions associated motor speech-related processes.³⁵

Section summary and definitions

Evidence from various task-based functional imaging paradigms in healthy people, direct cortical stimulation in neurosurgical populations and post-stroke lesion and white matter connection patterns all point to two lateral precentral gyrus speech areas, one ventral and one more dorsal. The existence of two speech-related zones has been appreciated for a while, but more recent neurosurgical and stroke-based connectivity studies (reviewed above) have clarified that neither of the two areas are centred on classical Broca’s area and instead are situated more on the precentral gyrus. Fig. 3 shows the location of the two regions, which we will call the ‘dorsal precentral speech area’ (dPCSA) and the ‘ventral precentral speech area’ (vPCSA)—localized from functional imaging and stimulation mapping studies—and shown in relation to cytoarchitectonic zones (solid colors) and relevant functional areas (outlines). By using these terms, we are not claiming speech selectivity (indeed, we believe the dPCSA is also strongly involved in song); rather, the terms merely highlight their prominent role in speech coordination. The dPCSA aligns with premotor cortex on the precentral gyrus (cytoarchitectonic area 6) while the vPCSA is mostly posterior to area 44 in a region that is also likely premotor cortex,⁴⁶ but does not yet appear in the Jülich-Brain probabilistic atlas. Note that the d/vPCSAs are predominantly on the crown of the gyrus and not in central sulcus where primary motor cortex is located. Thus, we also assume that each PCSA is an intermediate zone between lower-level primary motor systems and higher-level linguistic/cognitive systems (see ‘The dual speech coordination model’ section). When referring to the precentral speech areas themselves, we will use the acronyms dPCSA and vPCSA; when referring more broadly to the networks in which they are situated we will use descriptors such as ‘dorsal/ventral speech coordination networks’ or the like.

Functional differentiation of the dorsal and ventral precentral speech areas

In this section, we review three relatively recent findings that provide important clues to the unique role of the two speech coordination areas and which motivate a new hypothesis: that the dPCSA coordinates pitch-related vocalization (prosody/song) while the vPCSA coordinates phonetic/syllabic speech features.

The dorsal precentral speech area codes voice pitch during vocalization

One set of observations comes from an intracranial recording and stimulation study aimed at mapping motor cortex related to vocal pitch control.³¹ The primary task involved producing the same sentence repeatedly but with emphasis (higher pitch) on different words (I never SAID she stole my money; I never said SHE stole my money). After controlling for supralaryngeal articulation and intensity, a region on the mid lateral precentral gyrus was identified that showed neural activity time locked with the pitch accents. When the same sentences were played back to the participant, pitch-correlated auditory responses were recorded in the same region. A similar result was obtained with a non-speech singing task, showing that the effect is not speech-specific, consistent with a previous functional MRI (fMRI) study of speech and melody repetition.²⁹ Direct cortical stimulation of this zone produced involuntary vocalization. More ventral precentral cortical regions were not strongly pitch-tuned and instead were shown to be more related to control of voicing (which involves different muscles). The authors claim to have identified dorsal laryngeal motor cortex (dLMC), which serves to control vocal pitch. While we agree with their functional conclusion regarding this region’s role in pitch control, it is debatable whether it is dLMC or dPCSA. Specifically, because recordings were made on the gyral surfaces using implanted electrode grids, the investigators were more likely

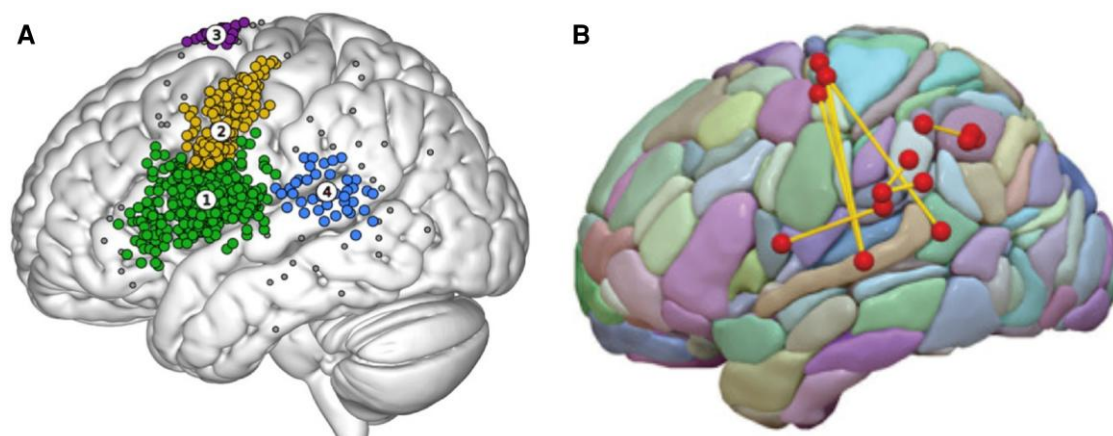


Figure 2 Neurosurgical and stroke-based evidence for dorsal and ventral speech coordination systems. (A) Map of locations eliciting speech arrest (see text) during direct cortical stimulation. Colours and numbers represent the results of a cluster analysis. The yellow (2) and green (1) clusters have density distributions centred on the dorsal and ventral precentral gyrus, respectively. From Lu et al.³⁰ (B) Map of white matter connections predictive of speech repetition deficits in post-stroke aphasia implicating both dorsal and ventral connectivity. From Baboyan et al.³⁸

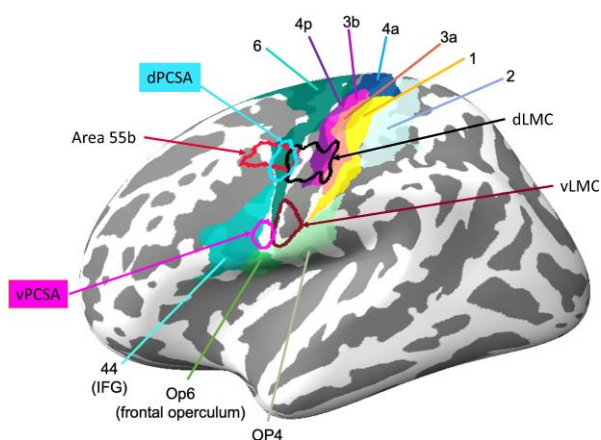


Figure 3 Surface-rendered map of functional and anatomic regions. Inflated brain map showing the dPCSA and vPCSA regions of interest in relation to the dorsal laryngeal motor cortex (dLMC), ventral laryngeal motor cortex (vLMC), area 55b, and maximal probability maps for cytoarchitectonic areas (colour-shaded areas) from the Jülich-Brain atlas,⁴⁷ supplemented by a thresholded probability map of premotor area 6.⁴⁸ dPCSA and vPCSA regions of interest are estimated from functional imaging activation peaks on the precentral gyrus in Rong et al.,¹⁷ which align with coordinates from the speech arrest stimulation study³⁰ (see Supplementary Fig. 1). The dLMC and vLMC were defined by performing an activation likelihood estimate (ALE) meta-analysis⁴⁹ of functional MRI studies that explicitly refer to laryngeal motor cortex in their results (see Supplementary Table 1). ALE values were thresholded at $P < 0.00001$ (cluster family-wise error correction at $P < 0.00001$ using 4000 permutations) and the resulting clusters were projected onto the fsaverage surface along with the vPCSA and dPCSA regions of interest. Area 55b was extracted from the Human Connectome Project's (HCP) Multimodal Parcellation 1.0 atlas.³⁷ Maximal probability maps for motor areas 4a and 4p, somatosensory areas 3a, 3b, 1 and 2, opercular areas Op6 and OP4, and area 44 were all extracted from the Jülich-Brain atlas⁴⁷ (version 2.9). The probability map for premotor area 6,⁴⁸ which is largely unmapped in the Jülich-Brain atlas, was projected onto the fsaverage surface, then thresholded to show vertices that were at least 40% likely to be area 6. See Supplementary material for more information.

recording from premotor rather than primary motor cortex, the latter being mostly buried in the central sulcus (Fig. 3) and where dLMC is argued to be located.⁵⁰ Thus, we suggest instead that the authors have uncovered some critical properties of the dPCSA.

The dorsal precentral speech area codes pitch features during perception

Further evidence for pitch-related sensitivity of the dPCSA comes from a completely different angle, an fMRI study aimed at mapping the speech-related spectrotemporal receptive field (STRF) properties of regions outside of auditory cortex.⁵¹ The study involved participants listening to speech that was filtered so that on different trials, portions of the spectrotemporal modulation spectrum were excised from the signal.⁵² Neural activity was then correlated with the presence or absence of signal in different regions of spectrotemporal modulation space enabling the estimation of regional STRFs.⁵³ Relevant to the present discussion, robust speech-related STRFs were found in a dorsal precentral location that aligns spatially with the dPCSA (Fig. 4). The vPCSA was not among the regions with mappable STRFs, although a more anterior site in the pars triangularis region of Broca's area did prove partly auditory responsive. The dorsal precentral site exhibited an STRF profile similar to that of early auditory cortex in that it showed tuning not only to regions of spectrotemporal modulation space relevant to intelligibility (typical of higher-order auditory cortex) but also to regions related to voice pitch ('atypical' of higher-order auditory cortex). Further, the dorsal precentral region showed strong functional connectivity with early auditory cortex, converging with the stroke connectivity result noted above. A recent intracranial cortical recording study provides evidence that auditory information engaging this dorsal precentral region has functional relevance for speech coordination: under conditions of fluency-disrupting delayed auditory speech feedback, activity in the dorsal precentral region was enhanced as part of a network that included auditory areas and the supramarginal gyrus.⁵⁴

In contrast, the STRF profile in the Broca's area site was selectively responsive to intelligible speech, was not tuned to pitch, and showed functional connectivity with higher-order regions in the superior temporal sulcus and middle temporal gyrus, not early auditory cortex, strongly suggesting that it is receiving inputs from higher-level (e.g. lexical) networks.

Persistent developmental stuttering implicates ventral sensorimotor cortex

Developmental stuttering persists into adulthood in approximately 1% of the population^{57,58} and results in sound and syllable

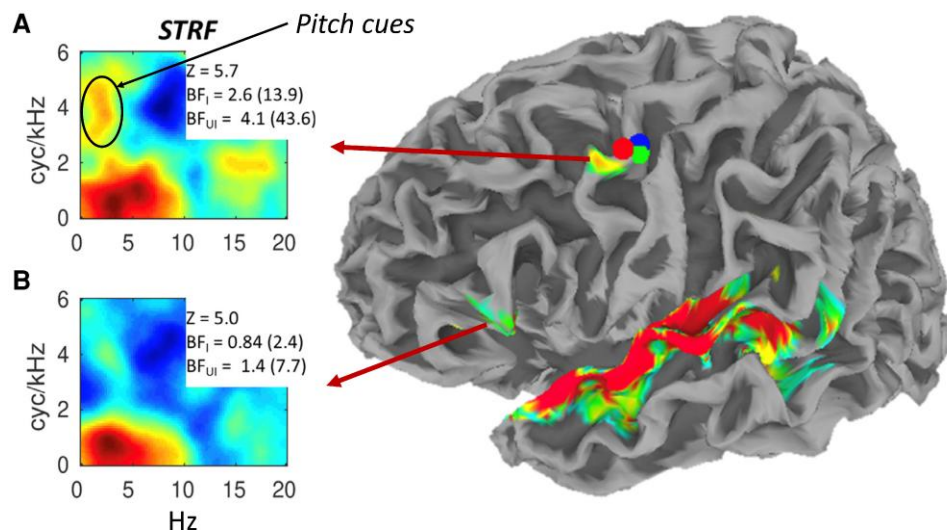


Figure 4 Spectrotemporal receptive field maps. Distribution of mappable STRFs (brain image) and plots of the STRF maps associated with the two lateral frontal STRF regions (left, A and B), one in dorsal precentral gyrus and one on the pars triangularis. The dorsal precentral site exhibits significant tuning to the pitch-related region of STRF space (black oval). Coloured dots correspond to the coordinates of three separate fMRI study activations, two which map auditory–motor speech responses (the slightly more posterior green and blue spheres^{28,55}) and one that mapped laryngeal motor cortex⁵⁶ with the coordinate on ‘gyral portion’ (see text). Figure from Venezia et al.⁵¹

repetitions, sound prolongations, and blocks, suggesting a deficit in the ability to initiate, control, and terminate speech gestures.⁵⁹ Fluency among adults who stutter has been reported to improve substantially during singing,^{60,61} suggesting a relative sparing of the more dorsal speech coordination system while disrupting the ventral speech coordination system at some level. This question was examined in a recent study of white matter structural connectivity of the dorsal and ventral speech systems using seeds for the dorsal and ventral laryngeal motor areas, respectively⁶² (see ‘Dorsal laryngeal motor cortex is the lowest level in the cortical hierarchy for vocal pitch-related coordination’ section for discussion of these two laryngeal areas). The ventral network was found to have weaker connectivity than the dorsal network in adults who stutter compared to controls and the connectivity profile of the ventral (but not dorsal) network predicted stuttering severity. This work suggests that the ventral network plays a critical role in speech coordination at the sound and syllable level and can be impaired while sparing song. It should be noted that prosodic abnormalities have also been reported in stuttering,⁶³ but relatively preserved singing ability in adults who stutter argue against a primary deficit in pitch-related voice coordination.

Section summary

The studies described above indicate that the dPCSA is involved both in the control of voice pitch and is capable of coding pitch-related cues in acoustic signals through its functional connectivity with relatively early auditory areas. This contrasts with the vPCSA, which appears to lack these pitch-related properties and instead is more involved in speech coordination at the phonetic and syllable level.

The dual-speech coordination model

We propose that dPCSA and vPCSA are part of distinguishable, parallel (but interacting) networks for coordinating pitch-related and syllabic/phonetic-related speech features, respectively. Specifically,

- The dPCSA is part of an auditory-weighted sensorimotor control circuit for pitch-related vocalization, which includes pitch-related prosodic aspects of speech as well as song, and predominantly acts on the laryngeal effector via the dorsal laryngeal motor cortex.
- The vPCSA is part of a somatosensory-weighted sensorimotor control circuit for syllabic and phonetic-related speech features and acts on supralaryngeal effectors as well as voicing-related control of the larynx via the ventral laryngeal motor cortex.

In the following sections we discuss the broader networks in which the two speech coordination areas are situated.

Parallel frontal motor hierarchies for speech coordination

Motor systems are hierarchically organized⁶⁴ and there is no reason to think that speech coordination systems are any different. What is less clear is whether the dPCSA and vPCSA are at a similar level in parallel hierarchical streams, as we will argue here, or whether one is subordinate to the other. (A third possibility is that the dorsal and ventral precentral speech regions are one undifferentiated zone,⁶⁵ the view we are challenging here.) One source of evidence for the parallel hierarchical claim is that both speech areas appear to be at the same level of the general cortical hierarchy somewhere between higher-order linguistic processing and lower-level motor implementation, as suggested by the cortical stimulation mapping data discussed earlier. Another source of evidence is the apparently tight relation between the dPCSA and dLMC, which we turn to next.

Dorsal laryngeal motor cortex is the lowest level in the cortical hierarchy for vocal pitch-related coordination

Humans have two spatially separable regions related to laryngeal control, one more ventral and thought to be in ‘premotor’ cortex, which is shared with other primates, and one more dorsal and thought to be in ‘primary’ motor cortex in the central sulcus, which is phylogenetically younger, not shared with other primates and

argued to be directly connected to brainstem motor nuclei controlling the larynx (unlike the ventral laryngeal area).^{50,65–67} We will refer to these regions as dLMC and ventral laryngeal motor cortex (vLMC). Functional differences between the dLMC and vLMC are yet to be mapped in detail, but one major distinction is the differential effects of cortical stimulation mapping: stimulating regions in the vicinity of dLMC elicits vowel-like phonation whereas more ventral stimulation does not, eliciting grunt-like behaviours instead.^{68,69}

While the existence of a more dorsal area involved in laryngeal control in addition to the more ventral area is unique to humans among primates, multiple motor maps for a given effector (e.g. the hand) has been observed in many species even within primary motor cortex and is thought to reflect an organization of motor cortex around adaptive behaviours or functional domains—such as reaching, grasping, hand-to-mouth, defensive postures and so on—rather than a simple somatotopy.^{70–72} Given that the same effector can be involved in multiple functional domains, multiple representations of the same effector is expected, as the data indeed show.⁷³ Applying this thinking to the larynx, the existence of two areas in humans suggests two functional domains associated with laryngeal control. The existing evidence strongly points to the dLMC as part of a functional domain for pitch-related vocalization and which includes other motor systems that are important for this functional domain, such as respiratory control.^{74–76} The functional domain(s) of the vLMC is less clear for speech, although some role in initiation, timing, and cessation may be involved.⁶² In non-human primates, the vLMC likely plays a role in non-vocal functions⁷⁷ such as ingestion.^{78,79} The broader ventral orofacial motor system, in which the monkey vLMC is situated, involves functions including oral grasping^{80,81} and communicative oral gestures such as lip smacking.⁸² It should be noted that limited voluntary control over innate vocalizations has been reported in monkeys, with ventral motor-related cortex implicated,^{83,84} consistent with a possible role of human vLMC in voicing initiation. However, damage to monkey ventral motor regions does not affect the ability to generate innate calls,⁸⁵ suggesting that the ventral orofacial motor system does not contain the neural codes for monkey vocalizations but rather a mechanism for voluntarily eliciting the vocalization ‘subroutines’ coded elsewhere.⁸⁶

The proposed framework is consistent with the hypothesis that vocal pitch control is the functional domain associated with dLMC. We make this inference on the basis of the spatial relation between the pitch-related dPCSA and the dLMC, which partially overlap in group-based studies (Fig. 3). In fact, some investigators refer to two subdivisions of the dLMC, sulcal and gyral LMC.⁸⁷ The gyral LMC partially overlaps our dPCSA, which we propose to be hierarchically dominant to the sulcal dLMC based on (i) the probabilistic location of dPCSA in premotor cortex and sulcal LMC in primary motor cortex (Fig. 3); (ii) dPCSA activation for imagined speech production (suggestive of a planning process)^{17,29}; and (iii) the acoustic response properties of the dPCSA in the absence of any expressive speech task,⁵¹ which is consistent with the previous observation of an anterior (auditory-motor) to posterior (more purely motor) gradient in this region (Fig. 1C).²⁸ Note that although we are framing our hypotheses around a rather strict hierarchy, it is likely that the boundaries reflect a gradient of hierarchical control rather than a sharply delineated hierarchy.

A higher-order language area anterior to the dorsal precentral speech area in the posterior middle frontal gyrus

Anterior to the dPCSA lies a higher-level language region in the posterior middle frontal gyrus, area 55b (Figs 1D and 3).³⁷ It is relatively unmyelinated and sandwiched between the two moderately myelinated lateral eye fields. Posterior to 55b is the heavily myelinated primary motor cortex,³⁷ which includes the (sulcal) dLMC. The dPCSA overlaps the posterior portion of area 55b, while the anterior portion of 55b extends into the posterior middle frontal gyrus. Evidence for a language function associated with anterior 55b comes from the very definition of the area, which is derived in part from the Human Connectome Project’s functional activation language task, story listening, which shows a peak activation focus in the anterior portion of the 55b parcellation.³⁷ Further evidence comes from a recent study showing a role for the posterior middle frontal gyrus in expressive agrammatism (an acquired deficit of syntactic production)⁸⁸ as well as from an automated meta-analysis of ‘syntactic’ processing that we conducted using the Neurosynth database⁸⁹ (Fig. 5).

What function(s) might be served by the higher-order level of the dorsal speech coordination network? This is an open question, but one hypothesis that is consistent with existing data is that it plays a role in prosodic planning, which is known to correlate with syntactic features and has been proposed as a planning frame for speech production.⁹⁰ One may wonder whether dysfunction of a prosodic planning system can impact expressive syntactic processing sufficiently to contribute to expressive agrammatism, as described above, a rather severe deficit of syntax. However, if ‘the prosodic structure of an utterance provides the representational “spine” that governs... the serial ordering of lexical elements and their sub-constituents’,⁹¹ (p. 11) then it seems quite possible that the disruption of this planning ‘spine’ could lead to significant deficits in the serial ordering of syntactic elements.¹⁰ Clinically, this hypothesis raises some potentially important questions regarding the role of prosodic deficits in aphasic symptomatology and the use of prosodic/intonational therapies for their treatment.^{92,93} For example, prosody-based therapies may only be useful in patients whose primary deficit does ‘not’ involve prosodic dysfunction, allowing such patients to make full use of prosodic information.⁹² That said, it is important to note that prosody is an umbrella term for a complex system involving at least two functional modes (linguistic and affective) and control over many acoustic features beyond pitch including loudness, duration, rhythm, and others.⁹⁴ Our conjecture that the higher-order posterior middle frontal gyrus region in the dorsal speech coordination hierarchy is related to prosodic planning does not specify which aspects of prosody may be involved, although given that pitch features are coded in the dPCSA, we would expect pitch-related prosodic information to be strongly involved. It is also unclear how selective this system is to linguistic prosody and whether there are hemispheric differences related to its function or relative contribution to linguistic versus affective prosody. We discuss these issues further in ‘Summary and outstanding questions’. For now, our point is that the posterior portion of the middle frontal gyrus, including anterior area 55b, appears to play some role in language planning as part of the dorsal hierarchy, apparently involving aspects of syntax, which we suggest may be related to prosodic planning of some sort.

The ventral speech coordination hierarchy

The proposed hierarchy in the dorsal system parallels a similar hierarchy in more classic ventral speech and language areas with the vPCSA lying between primary oromotor cortex posteriorly and Broca’s area anteriorly in the posterior inferior frontal gyrus (Fig. 3). We have argued that the vPCSA coordinates phonetic

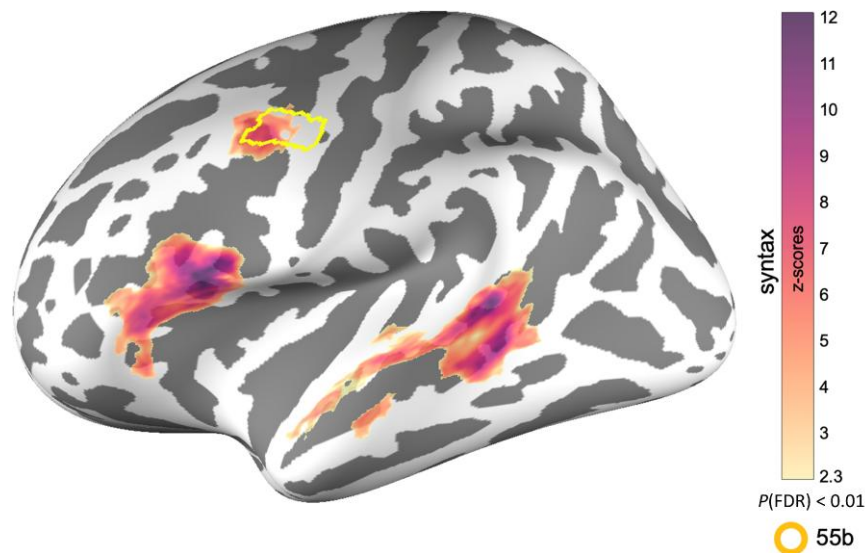


Figure 5 An automated meta-analysis of functional imaging studies on the term ‘syntactic’ (heat map) using neurosynth.org. Note the relation to area 55b (yellow outline). See [Supplementary material](#) for more information about methods and analysis.

articulation within the syllabic frame^{95,96} and we follow previous hypotheses that higher levels of the ventral speech coordination system play a role in sequencing of syllables (pars opercularis)^{17,96} and morphosyntactic elements (pars triangularis)¹⁰ (see also Hagoort⁹⁷). As the existence of a ventral frontal speech- and language-related hierarchy is widely accepted, we will not further review or argue the details of its internal organization, which is an active topic of research.

Connectivity

Results from the connectome mapping study of repetition deficits in stroke aphasia (see section above, ‘Two speech areas in the lateral precentral gyrus’; [Fig. 2B](#)) indicate different patterns of connectivity associated with the dPCSA and vPCSA with the dorsal region more strongly connected to auditory cortex and the ventral region more strongly connected to inferior parietal/secondary somatosensory cortex. The spectrotemporal receptive field mapping study described in the ‘Functional differentiation of the dorsal and ventral precentral speech areas’ section also included a functional connectivity analysis that showed connectivity between the dPCSA and auditory cortex. Here we present two new analyses of functional connectivity of the two speech areas. Both use seed regions of interest corresponding to the dPCSA and vPCSA defined from Rong et al.¹⁷ ([Fig. 3](#) and [Supplementary Fig. 1](#)).

One functional connectivity analysis is from a task fMRI study of speech-based spectrotemporal receptive field mapping⁹⁸ using similar methods to the study described in ‘Functional differentiation of the dorsal and ventral precentral speech areas’. The other functional connectivity analysis draws on previously published multi-echo resting-state fMRI data.⁹⁹ A portion of this data set is publicly available.¹⁰⁰ The results of both functional connectivity analyses are presented in [Fig. 6](#), which confirms the hypothesized pattern of stronger auditory cortex connectivity for the dPCSA and stronger higher-level somatosensory connectivity for the vPCSA.

Section summary

This section argued for parallel hierarchical networks supporting speech coordination. Each involves a primary motor cortical zone

(or zones) at the lowest level, an intermediate level corresponding to the two precentral speech areas, and one or more higher-order linguistic region. The ventral hierarchy corresponds to the familiar inferior frontal speech/language network with Broca’s area at the top of the hierarchy, with an internal hierarchical organization of its own. The dorsal hierarchy involves dLMC, the dPCSA, and the posterior middle frontal gyrus language area, which we hypothesize is a core region involved in prosodic planning. The precentral speech areas differ in terms of the dominant pattern of sensory input they receive with the dPCSA more strongly auditory-weighted and the vPCSA more strongly somatosensory-weighted. [Fig. 7](#) provides a graphical sketch of the proposed architecture.

We have proposed a new architecture for speech coordination, which divides the system into two parallel but interacting hierarchies. While we believe the arguments presented above stand on their own, given the novelty of the proposed architecture, it is helpful to consider ‘why’ such an architecture may have evolved. If there are good reasons to believe the proposed architecture makes sense evolutionarily, this would increase the plausibility of our neural architecture claims. Moreover, because coordination of pitch and articulate speech are highly integrated in the fully evolved human system, consideration of separable evolutionary histories for these abilities may provide important clues about how to best disentangle the independent contributions of dorsal and ventral speech coordination systems, which could yield important clinical insights as suggested above. We turn to this question next.

Part 2: evolution of the dual-speech coordination architecture

The evolutionary origin of human speech and language ability has been a topic of deep interest and much debate since Darwin, who conjectured that a song-based communication system evolved as a precursor to speech.^{104,105} This idea, which has modern-day supporters,^{106,107} may hold some relation to dual-speech coordination systems in that the presently proposed neural architecture model argues for separable mechanisms controlling song and articulate speech, raising the possibility of separable evolutionary histories.

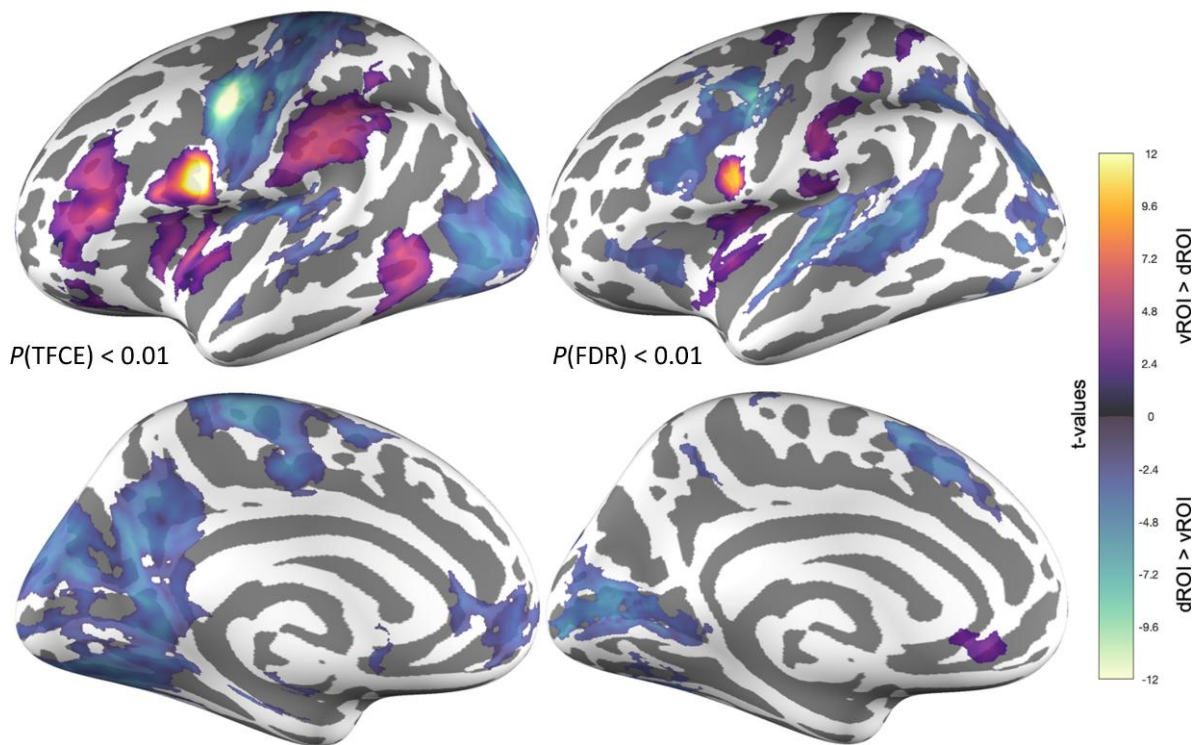


Figure 6 Functional connectivity analyses for dorsal versus ventral precentral speech areas. (Left) Resting state functional connectivity map for the dPCSA (cool colours) and vPCSA (warm colours). Whole brain maps were generated for each region of interest (defined in Fig. 3) in each participant ($n = 137$) using multi-echo independent components regression.¹⁰¹ Voxelwise paired t -tests were then performed to determine brain regions where resting functional connectivity was significantly higher to either dPCSA or vPCSA in the group. Family-wise error was controlled at $P < 0.01$ using threshold free cluster enhancement (TFCE) with 5000 permutations.¹⁰² The resulting t -map was projected onto the fsaverage surface using registration fusion with Advanced Normalization Tools.¹⁰³ (Right) Task fMRI-based, functional connectivity maps. Beta time series functional connectivity analyses ($n = 34$) were conducted using data from 400 trials of a receptive competing speech task⁹⁸ ('Competing' condition). Data were projected onto a cortical surface model of the MNI152 template and surface-node-wise estimates of functional connectivity (least-squares-separate beta time series; Pearson correlation) were obtained using the average time series within dPCSA and vPCSA regions of interest defined as in Fig. 3 as seeds. A second-level contrast t -map (dPCSA connectivity versus vPCSA connectivity) was obtained and thresholded at FDR-corrected $P < 0.01$, then projected onto the fsaverage surface using registration fusion with Advanced Normalization Tools.¹⁰³ See [Supplementary material](#) for more information about methods and analysis.

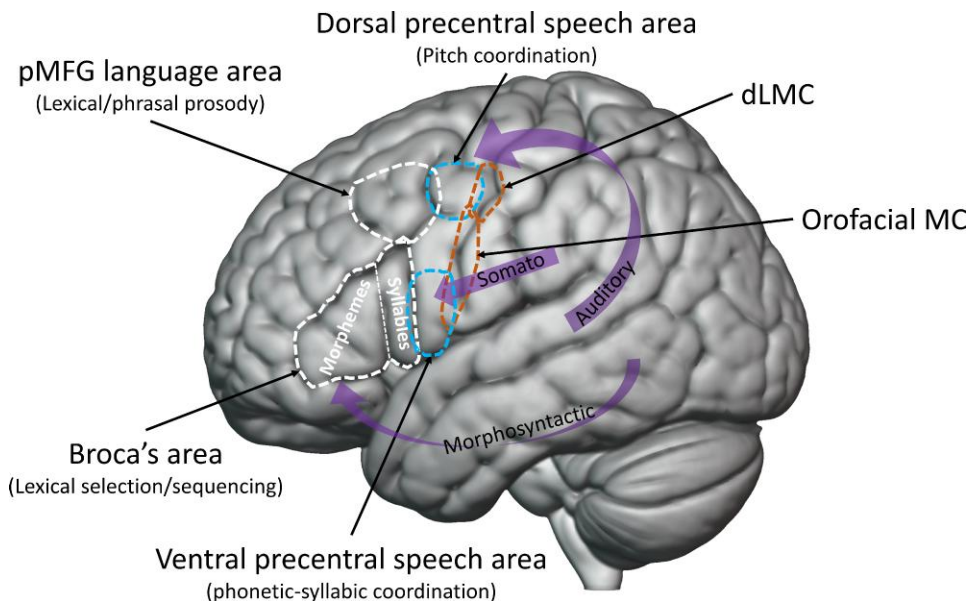


Figure 7 Schematic depiction of the dual speech coordination system model. Input labels (arrows) are intended to reflect a bias toward the named input type, not exclusivity, and are not intended to reflect anatomical locations of white matter pathways. For detailed arguments regarding the morpho-syntactic pathway, see Matchin and Hickok.¹⁰ Anatomical localizations are approximate. Location of orofacial motor cortex ventral to dLMC is based on a meta-analysis from Guenther.⁵⁹

It is also relevant that the dPCSA receives auditory-biased sensory inputs and is positioned between cortical eye fields, which, also make use of spectral acoustic cues (see next section, ‘The dPCSA evolved from a premotor goal-directed orienting system’). In what follows we evaluate the possibility that the dPCSA evolved in its current location because the dorsal premotor cortex had the right kinds of sensory inputs in our pre-speech ancestors. We also review existing claims regarding the evolution of the ventral system for supralaryngeal articulatory coordination, followed by a discussion of the potential sequence of emergence of dorsal and ventral speech coordination systems, and then a brief consideration of the relation between our proposal and theories of the evolution of language.

The dorsal precentral speech area evolved from a premotor goal-directed orienting system

We propose that there is a relation between auditory inputs to premotor cortex that are used in the service of goal-directed orienting and the evolution of the dorsal speech coordination system (see Martins and Boeckx¹⁰⁸ for more general discussion of the evolutionary relation between language and attention systems). The logic of this argument is that (i) motor coordination systems require the right kinds of sensory inputs to enable feedback control^{59,109}; (ii) control of vocal pitch features requires input from auditory pitch-related codes; and (iii) the only cortical motor network in the non-human primate brain known to use pitch-related codes is the orienting system, which involves the dorsal precentral region. Therefore, if a motor coordination system for pitch-related vocalization were to evolve, it should be in the vicinity of networks that already have the relevant sensory input signals, which is the dorsal premotor cortex. We elaborate this argument next.

Spatial orienting systems and the dorsal precentral speech area are adjacent

In the macaque, the dorsolateral premotor cortex and frontal eye fields are part of the dorsal sensorimotor stream and receives input from caudal non-primary auditory areas that are thought to play a role in spatial localization of auditory signals.^{110,111} The frontal eye field in particular has a large population of auditory responsive cells¹¹² and is well-suited to guiding sensory-driven orienting responses and covert spatial attention.¹¹³ The fact that the frontal eye field utilizes not only visual but also auditory inputs and is part of the orienting system is key given that human area 55b and the dPCSA are situated in between the human frontal eye field dorsally and the premotor eye field ventrally.³⁷ Human fMRI work has also identified auditory-driven spatial attention networks in the dorsolateral premotor cortex.^{114,115}

Spatial orienting uses spectral auditory cues including pitch

While auditory spatial localization in the horizontal plane relies primarily on sound level and time delay differences in acoustic signals reaching the two ears, localizing sounds in the vertical plane relies strongly on spectral acoustic features,¹¹⁶ including pitch.¹¹⁷ Natural auditory scene statistics exhibit an elevation-based frequency gradient with higher-pitched sounds tending to have sources at higher elevations.¹¹⁸ Human listeners use these cues to modulate spatial orienting and inform sound source elevation judgements.^{117–120} Use of spectral cues in spatial orienting must be evolutionarily adaptive given that the filtering properties of the human outer ear have evolved to amplify this vertical pitch

gradient.¹¹⁸ The ability to detect these cross-modal correspondences between auditory pitch and visuospatial elevation has also been reported in human neonates¹²¹ as well as in domestic dogs,¹²² arguing against a cultural explanation. More generally, given that vocalized spectral cues are correlated with animal size¹²³ and that modulations of these cues are used for social communication in a variety of mammals,¹²⁴ it should be no surprise that such information is salient to listeners and used by the orienting/attention system¹²⁵ beyond its relevance as a cue to spatial location.¹²⁶ Relatedly, the human frontal eye field exhibits ‘ultra-rapid’ responses to both auditory and visual stimuli, which are modulated by sensory features, including pitch, as opposed to simply the presence or absence of stimulation.¹²⁷ These ultra-rapid responses are likely useful for rapid orienting and suggest that relatively low-level auditory features are transmitted in this signal, as opposed to higher-level object-based features, which would require more processing and thus longer latencies.

Auditory-responsive regions have also been identified in more ventral prefrontal^{110,128} and premotor cortex^{129,130} in monkeys. However, in the prefrontal case, inputs come from rostral auditory cortex, are part of the ventral ‘what’ stream, and code more complex information. Audio-motor responses have been recorded in this region, involving the execution and perception of species-specific calls.¹³⁰ These responses, however, likely reflect the ability to modulate innate calls in response to external stimulation rather than using auditory features to code the calls themselves.^{130,131} In the premotor case, one study reported neurons that are spatially selective for proximal locations near the body, but which are not tuned to pitch, instead preferring broadband sounds, and with somatosensory receptive fields around the ears or back of the head, ideal for guiding actions toward objects (e.g. insects) around the head.¹²⁹

Thus, dorsal premotor cortex and nearby eye fields, by virtue of the auditory orienting networks located there, but not ventral premotor cortex, was arguably prewired in our ancestral species to receive a rich array of acoustic information, including pitch-related cues, which would have been critical for an evolving auditory-motor, pitch-related vocal production-learning circuit. Comparative evidence from the evolution of song in birds indicates that brain areas with pre-song auditory inputs are targets for the evolution of song.¹³²

The premotor cortex as a hub for a broader goal-directed orienting system

Auditory cues can drive goal-oriented action beyond eye-based spatial localization and again dorsolateral premotor areas are implicated in this ability in human and non-human primates. For example, some studies have reported that auditory rhythmic entrainment engages dorsal premotor cortex while visually paced entrainment does not, the latter involving more ventral premotor regions.^{133–135} Instrumental musicians, who have stronger auditory–manual associations, exhibit stronger activation in dorsal premotor cortex than non-musicians.¹³⁶ Monkeys, when trained to reproduce pitch sequences via a ‘monkey piano’, tend to activate more dorsal motor and premotor arm-related areas while passively listening to the auditory sequences they previously learned to ‘play’.¹³⁷

Research on the frontal eye fields also points to a broader orienting and attention network.¹³⁸ Indeed, while the name ‘eye field’ implies eye-selective activity, research in human and non-human primates that examined both eye and head orienting behaviour has found evidence that the frontal ‘eye’ field is involved in control

and coordination of both effectors,^{139,140} suggesting that these fields are part of a hub for a broader orienting network.^{141,142} Thus, we propose that it is out of this dorsal premotor, auditory-responsive, goal-directed orienting system that voluntary laryngeal control and the dPCSA evolved.

Behavioural evidence for a link between vocal pitch control and non-vocal orienting

If the dPCSA and its associated network evolved from a more general orienting system, we might expect to see correlations between vocal and non-vocal gestures due to coactivation of related networks. Such correlations have been observed both in musical and linguistic contexts.

Behavioural studies have found a strong correlation between laryngeal pitch control and head/face movements. Singers, for example, raise their head and eyebrows with an amplitude that strongly correlates with the size of ascending pitch intervals.¹⁴³ This cannot be explained as a strategy for improving singing quality as head movement is not advantageous for pitch-related vocal production.¹⁴⁴ Head movements during talking are also correlated with prosodic pitch¹⁴⁵ and lexical tone.¹⁴⁶ These non-vocal head and eye movements are correlated strongly enough with pitch to provide observers with redundant visual cues to pitch information. Pitch accents in speech are also correlated with co-speech manual gestures¹⁴⁷ and may be causally related: two expressions of the same prosodic frame for speech planning.⁹¹ All of this points to a correspondence between laryngeal pitch control and non-vocal gesturing, which is consistent with the hypothesis of a phylogenetic connection between them.

Coordination of the supralaryngeal vocal tract evolved from ventral orofacial control systems

Frame-content theory

MacNeilage's frame-content theory of the evolution of speech production provides an account of how the ability to articulate phonological segments within syllabic frames might have evolved.^{95,96} The basic idea is that syllables are formed by the cyclic opening and closing of the jaw, into which various phonetic gestures are articulated to form different vowels (open phase) and consonants (closed phase). Syllabic jaw cycles, according to the frame-content theory, evolved from cyclic jaw actions characteristic of mastication and observed in communicative lip smacking in non-human primates. Lip smacking in particular, while unaccompanied by vocalization (but see below), involves speech-like gestures including synchronized lip, jaw and tongue movements with jaw oscillation cycles in the speech frequency range.¹⁴⁸ Thus, lip smacking is a very strong candidate for a pre-adaptation preceding supralaryngeal speech coordination.^{95,96,149}

Lip smacking implicates ventral frontal premotor regions

Although MacNeilage⁹⁵ argued for the medial motor system (e.g. the supplementary motor area, SMA) as the neural source of cyclic jaw movement, recent work in macaques has pointed to the lateral ventral motor network for both ingestive actions and lip smacking behaviour. Ferrari et al.,⁸⁰ for example, identified neurons in F5, a presumed homologue of human Broca's area, that responded during a range of executed mouth actions including grasping, sucking, chewing, and lip smacking (among others). Ingestion-related action responses were the most commonly recorded. About half (48.7%) of

the neurons also had visual-response properties. Auditory responses were not studied. A more recent fMRI study of macaques producing lip smacking gestures reported activation in a broad motor network including primary motor cortex and ventral (but not dorsal) premotor cortex, which overlapped that for non-communicative facial movement control.⁸² This strongly suggests that lip-smacking communicative gestures developed from ventral orofacial motor control networks.

The ventral motor regions that are activated during the generation of lip-smacking actions do not seem to respond to social communicative auditory signals. In one unit-recording study, macaque F5 neurons were reported to respond to object-action-related sounds such as paper being torn, but no responses to monkey calls were found.¹⁵⁰ In another study,⁸³ macaques were trained to elicit 'coo' calls in response to the presentation of food. A population of F5 cells responded prior to or around the onset of cued calls, but they did not respond to auditory recordings of the calls or during spontaneous call generation (consistent with the view that ventral motor areas do not house the codes for innate calls but play a role in voluntarily initiating them).

These studies provide evidence for ventral orofacial motor networks as the basis for the evolution of supralaryngeal vocal tract coordination and reinforce the view that this system was not particularly sensitive to vocal auditory signals in our pre-human ancestors.

Alternative hypotheses regarding the function and location of the dorsal and ventral systems

Brown and colleagues have proposed an alternative theory of the location of dLMC (and by extension the dPCSA), arguing that it is related to the more dorsal distribution of motor cortex for jaw opening.⁸⁷ However, a close examination of the distribution of jaw opening motor areas in the monkey do not provide strong support for this claim, with responses during jaw-open and jaw-close occurring both ventrally and more dorsally.⁷⁸ Belyk et al.⁶⁹ suggest that both dLMC and vLMC play a role in integrating respiration with phonation during vocalization, but hint that functional differences between vLMC and dLMC may reflect the fact that (i) only vLMC seems to respond to somatosensory stimulation of the larynx; and (ii) vLMC appears to be cytoarchitecturally intermediate to primary motor and primary somatosensory cortex. We view this hypothesis as broadly consistent with ours in that it supports the view that a primary organizing feature of dorsal and ventral speech coordination circuits is their relative sensitivity to different sensory inputs (i.e. auditory versus somatosensory), and argues against an organization based on motor somatotopy.

Evolutionary sequence of dorsal and ventral speech coordination systems

Given the arguments in earlier sections, it is reasonable to hypothesize that in our pre-speech/song ancestor, the ventral vocal tract/vLMC system coordinated orofacial actions for mastication, oral grasping, and communicative lip-smacking gestures. The dorsal system at this time had nothing to do with vocalization and instead supported orienting and attention. The next step in our evolution towards speech, we hypothesize, following Darwin and others,^{104,107} was the evolution of some form of pitch-based song, enabled by the evolution of the dLMC with its direct connectivity to the brain stem nucleus innervating the larynx.^{86,151} The dLMC inherited from the orienting network an auditory-weighted motor coordination circuit mediated by the dPCSA and perhaps

auditory-motor area Spt.¹⁵² Once a pitch-based song-like system was in place (at least in rudimentary form), the ventral system, which already had voluntary control over orofacial gestures, was refined for phonetic/syllabic coordination mediated by the emergence of the vPCSA and even higher-level coordination systems in Broca's area. In functional terms, we are claiming that pitch coordination preceded and is foundational to phonetic coordination.

Four pieces of evidence support the hypothesis that pitch coordination preceded phonetic coordination. The first is that pitch control is a simpler engineering task in that it can be modelled as a simple linear system using only auditory feedback; in contrast, controlling the multidimensional supralaryngeal system for speech articulation requires non-linear sensorimotor mappings and integration of somatosensory feedback.¹⁵³ Second, prosody, which prominently includes pitch melody and accent, plays an important role in word learning¹⁵⁴ and, as noted, functions as an important planning frame for speech. Thus, if pitch coordination evolved first, it could potentially facilitate the more computationally complex task of developing supralaryngeal vocal tract control. This leads to our third piece of evidence: pitch control in the form of song has evolved in multiple song species,¹⁵⁵ but the ability to articulate speech by coordinating both the sound source and filtering organs is rare, found only in parrots and humans,¹⁵⁶ both of which also have pitch control. Based on the first two points and given this pattern of vocal learning ability, it may be that song is a necessary pre-adaptation for fully coordinated speech. Fourth, the development of vocalization in infants progresses from cooing, a song-like vocalization, to a later syllable-like babbling stage.¹⁵⁷ Interestingly, direct cortical recordings of a 10-month-old during cooing and babbling vocalizations showed neural activation in distinct dorsal (cooing) and ventral (babbling) precentral gyrus locations.¹⁵⁸ All of these observations point to an earlier evolution of the dorsal pitch-related speech coordination system prior to the modern-day ventral speech coordination system.

Some readers may object that monkeys can, with much effort and explicit training, gain partial voluntary control over the elicitation of innate calls (see 'Coordination of the supralaryngeal vocal tract evolved from ventral orofacial control systems'). If simple vocalizations could be brought under full voluntary control to initiate or cease, then this would provide a vocal source signal enabling further evolution of auditory-guided coordination of the supralaryngeal filter. A pitch coordination system, on this scenario, is an add-on rather than a prerequisite. An apparent existence proof showing perhaps a first step in such an evolutionary trajectory comes from the vocalized lip smack or 'wobble' of the gelada (*Theropithecus gelada*), which appears to involve overlaying lip smacking on a 'moan' vocalization, both of which are separately present in their repertoire.¹⁵⁹ This does not explain the evolution of pitch coordination, but it does complicate the claim that the pitch coordination system evolved first. A second possible objection comes from vocal convergence in great apes (e.g. Watson et al.¹⁶⁰)—a simple form of vocal learning where vocalization of individuals in a group become similar—or the apparent ability of orangutans to modify the pitch (higher versus lower) of an existing call in response to a higher or lower pitch cue.¹⁶¹ Both kinds of results seem to challenge the view that vocal learning is unique to humans among primates. But these results themselves have been challenged on methodological grounds³⁶ and even if valid are not far outside the range of vocal control ability across many mammalian species, for example, in modulating spectral features, including pitch, under some conditions.

Finally, it is notable that a holistic characterization of speech prosody must consider pitch, timing, level, and timbre,¹⁶² the joint

control over which surely requires integration between the dorsal and ventral speech coordination systems. As we note below, a crucial piece of the evolutionary puzzle is not only 'why' these separate systems came into existence but 'how' they came to be integrated for coordination across multiple dimensions of speech motor control. The fine-grained details of this integration may explain how damage to one or the other system can be expected to impact complex speech behaviours; for example, whether damage to 'either' system produces wholesale disruption of behaviours that depend critically on their integration or, alternatively, whether individual behavioural elements can be selectively disrupted (e.g. elements of prosody such as pitch contour versus speech rate).¹⁶³ So, while it may be true that observations such as these complicate the claim that pitch-related laryngeal coordination pre-dates supralaryngeal coordination by raising the possibility that much can be accomplished without the dorsal pitch-related system, or that the dorsal system alone is not sufficient to support all elements of pitch and prosody as they are realized in a fully integrated system, neither do they provide compelling evidence against it.

Relation to theories of the evolution of language

As noted, the present dual coordination hypothesis is consistent with Darwin's song-first hypothesis¹⁰⁴ and indirectly supports it by showing that song and articulate speech involve distinguishable systems, the latter being more computationally complex. Other authors, however, have argued for an early mimetic or gestural protolanguage stage in the evolution of language.^{164–167} Although vocal and gestural foundations of language are often pitted against one another, both could serve foundational roles as part of an integrated system.^{168–170} Indeed, they appear to do so developmentally: both gestural and prosodic communication emerge early in human development and both are argued to play foundational roles in lexical, syntactic, and pragmatic aspects of language acquisition.¹⁷¹

The present dual coordination model builds a possible bridge between these evolutionary hypotheses in that it provides a plausible neural link between pitch-related vocal coordination and non-vocal gesture, both of which are associated with a dorsal, premotor orienting system as argued above.

Summary and outstanding questions

Many questions remain to be addressed. For example, coordination 'between' the two systems is clearly necessary and disruption of such coordination is surely a cause of speech coordination deficits such as apraxia of speech. This will be a critical topic for future work. Additionally, rhythm is an important feature in prosodic and song production and can be disrupted in speech coordination deficits. Given that rhythmic timing also plays an important role in phonetic planning, it will be crucial to determine whether and to what extent vocal rhythm production depends on dorsal versus ventral speech coordination circuits (or their interaction). Perhaps rhythmic synchronization, the ability to synchronize movement to an auditory beat—a rare trait found only in species with complex vocal learning—is a necessary function enabling the coordination of the two proposed streams.¹⁵⁶ Respiratory control is yet another important component of speech coordination that needs to be considered in relation to the two systems in future investigations.⁶⁹ There are also more detailed functional-anatomic questions to answer, such as the precise boundaries of these systems, the function of cortex in the inferior frontal sulcus, which lies between the two proposed hierarchies, and the details of hierarchical levels.

One may wonder whether previous work on the neurology of affective prosody, which is argued to be supported by the posterior ‘inferior’ frontal gyrus in the right hemisphere,¹⁷² contradicts the present hypothesis of pitch-related prosodic control by the dPCSA. We do not believe it does for two reasons. First, while it’s true that patients with expressive affective prosody deficits often have lesions involving the right inferior frontal gyrus, the degree of hemispheric specialization for affective prosody remains questionable¹⁷³ and there have been no large-scale lesion–symptom mapping studies to confirm the specificity of the inferior frontal gyrus claim in a large sample that can control for lesion size and *a priori* density distributions. Second, a recent analysis of the acoustic speech features that are disrupted in cases of expressive aprosodia pointed toward segment duration timing and voice timbre features rather than pitch height control.¹⁷⁴ Given that prosody is a complex system involving many features beyond pitch, it will be important in future studies to isolate pitch from other features.

Finally, there are important questions about lateralization that need to be answered. It is well-established that ventral coordination system is left dominant, but the dorsal system has not been studied extensively. Several recent direct cortical stimulation/recording studies suggest responses to acoustic speech features and laryngeal production in left ‘and’ right dPCSA/dLMC,^{31,175–177} and functional activation of the dPCSA in auditory-motor speech and song tasks is bilateral, unlike other auditory-motor areas such as Spt, which is left dominant.²⁹ This suggests that the dPCSA itself (functionally the coordination of vocal pitch) is organized more or less bilaterally. Higher levels of the dorsal system may exhibit a stronger lateralization pattern given that the association between syntactic function and the posterior middle frontal gyrus (e.g. Fig. 5) tend to be left dominant. The long-held view that the right hemisphere is dominant for affective prosody¹⁷² is also consistent with the existence of asymmetries at levels of speech production above those mediated by the precentral speech areas. However, while it’s true that one does commonly find affective aprosodia following right hemisphere damage, it appears that similar deficits are also seen following left hemisphere damage, questioning the strength of the asymmetry.¹⁷³ Similarly, research has failed to find strong lateralization effects for linguistic prosody.^{94,173} This leaves us in a rather murky state regarding the lateralization patterns of the dorsal stream in particular and especially where it applies to different hierarchical levels and to different subfunctions. Additionally, there is evidence for a possible hemispheric dissociation of song/prosody production versus production other aspects speech/language versus breathing control that remain to be sorted.^{76,178,179} Also of interest is the relation of these (a)symmetry patterns in speech production to hypothesized asymmetry patterns in auditory perception.^{180,181} We leave these and other questions to future study.

Acknowledgements

We’d like to thank Kourosh Saberi, Ravindra Arya, Oren Poliva, Tecumseh Fitch, and Cedric Boeckx for helpful comments. This material is the result of work partially supported with resources from and the use of facilities at the VA Loma Linda Healthcare System, Loma Linda, CA, USA. The contents do not represent the views of the U.S. Department of Veterans Affairs or the U.S. government.

Funding

This work was supported by NIH NIDCD R01 DC009659 (G.H.) and by the U.S. Department of Veterans Affairs, Veterans Health

Administration, Rehabilitation Research & Development Service Award IK2RX002702 to J.H.V.

Competing interests

The authors report no competing interests.

Supplementary material

Supplementary material is available at *Brain* online and data are available upon request.

References

1. Broca P. Remarques sur le siège de la faculté du langage articulé; Suivies d’une observation d’aphémie (perte de la parole). *Bull Soc Anat (Paris)*. 1861;6:330–357, 398–407.
2. Broca P. Sur le siège de la faculté du langage articulé. *Bull Soc Anthropol*. 1865;6:337–393.
3. Berker EA, Berker AH, Smith A. Translation of Broca’s 1865 report. Localization of speech in the third left frontal convolution. *Arch Neurol*. 1986;43:1065–1072.
4. Green CD. Remarks on the seat of the faculty of articulated language, following an observation of aphemia (loss of speech). <http://psychclassics.yorku.ca/Broca/aphemie-e.htm>
5. Geschwind N. Language and the brain. *Sci Am*. 1972;226:76–83.
6. Fedorenko E, Blank IA. Broca’s area is not a natural kind. *Trends Cogn Sci*. 2020;24:270–284.
7. Amunts K, Schleicher A, Burgel U, Mohlberg H, Uylings HB, Zilles K. Broca’s region revisited: Cytoarchitecture and inter-subject variability. *J Comp Neurol*. 1999;412:319–341.
8. Rogalsky C, Matchin W, Hickok G. Broca’s area, sentence comprehension, and working memory: An fMRI study. *Front Hum Neurosci*. 2008;2:14.
9. Conner CR, Kadipasaoglu CM, Shouval HZ, Hickok G, Tandon N. Network dynamics of Broca’s area during word selection. *PLoS One*. 2019;14:e0225756.
10. Matchin W, Hickok G. The cortical organization of syntax. *Cereb Cortex*. 2020;30:1481–1498.
11. Fedorenko E, Duncan J, Kanwisher N. Language-selective and domain-general regions lie side by side within Broca’s area. *Curr Biol*. 2012;22:2059–2062.
12. Grodzinsky Y, Santi A. The battle for Broca’s region. *Trends Cogn Sci*. 2008;12:474–480.
13. Schell M, Zaccarella E, Friederici AD. Differential cortical contribution of syntax and semantics: An fMRI study on two-word phrasal processing. *Cortex*. 2017;96:105–120.
14. Novick JM, Trueswell JC, Thompson-Schill SL. Cognitive control and parsing: Reexamining the role of Broca’s area in sentence comprehension. *Cogn Affect Behav Neurosci*. 2005;5: 263–281.
15. Buchsbaum BR, D’Esposito M. The search for the phonological store: From loop to convolution. *J Cogn Neurosci*. 2008;20: 762–778.
16. Peeva MG, Guenther FH, Tourville JA, et al. Distinct representations of phonemes, syllables, and supra-syllabic sequences in the speech production network. *Neuroimage*. 2010;50:626–638.
17. Rong F, Isenberg AL, Sun E, Hickok G. The neuroanatomy of speech sequencing at the syllable level. *PLoS One*. 2018;13: e0196381.
18. Fiebach CJ, Schubotz RI. Dynamic anticipatory processing of hierarchical sequential events: A common role for Broca’s

- area and ventral premotor cortex across domains? *Cortex*. 2006;42:499–502.
19. Strand EA, Duffy JR, Clark HM, Josephs K. The apraxia of speech rating scale: A tool for diagnosis and description of apraxia of speech. *J Commun Disord*. 2014;51:43–50.
 20. Dronkers NF. A new brain region for coordinating speech articulation. *Nature*. 1996;384:159–161.
 21. Ogar J, Willock S, Baldo J, Wilkins D, Ludy C, Dronkers N. Clinical and anatomical correlates of apraxia of speech. *Brain Lang*. 2006;97:343–350.
 22. Chang EF, Kurteff G, Andrews JP, et al. Pure apraxia of speech after resection based in the posterior middle frontal gyrus. *Neurosurgery*. 2020;87:E383–E389.
 23. Patira R, Ciniglia L, Calvert T, Altschuler EL. Pure apraxia of speech due to infarct in premotor cortex. *Neurol Neurochir Pol*. 2017;51:519–524.
 24. Itabashi R, Nishio Y, Kataoka Y, et al. Damage to the left precentral gyrus is associated with apraxia of speech in acute stroke. *Stroke*. 2016;47:31–36.
 25. Hickok G, Rogalsky C, Chen R, Herskovits EH, Townsley S, Hillis AE. Partially overlapping sensorimotor networks underlie speech praxis and verbal short-term memory: Evidence from apraxia of speech following acute stroke. *Front Hum Neurosci*. 2014;8:649.
 26. Graff-Radford J, Jones DT, Strand EA, Rabinstein AA, Duffy JR, Josephs KA. The neuroanatomy of pure apraxia of speech in stroke. *Brain Lang*. 2014;129:43–46.
 27. Price CJ, Wise RJS, Warburton EA, et al. Hearing and saying: The functional neuro-anatomy of auditory word processing. *Brain*. 1996;119:919–931.
 28. Wilson SM, Saygin AP, Sereno MI, Iacoboni M. Listening to speech activates motor areas involved in speech production. *Nat Neurosci*. 2004;7:701–702.
 29. Hickok G, Buchsbaum B, Humphries C, Muftuler T. Auditory-motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. *J Cogn Neurosci*. 2003;15:673–682.
 30. Lu J, Zhao Z, Zhang J, et al. Functional maps of direct electrical stimulation-induced speech arrest and anomia: A multicentre retrospective study. *Brain*. 2021;144:2541–2553.
 31. Dichter BK, Breshears JD, Leonard MK, Chang EF. The control of vocal pitch in human laryngeal motor cortex. *Cell*. 2018;174:21–31.e9.
 32. Broca P. Nouvelle observation d'aphémie produite par une lésion de la troisième circonvolution frontale. *Bull Soc Anat (Paris)*, 2e serie. 1861;6:398–407.
 33. Dronkers NF, Plaisant O, Iba-Zizen MT, Cabanis EA. Paul Broca's historic cases: High resolution MR imaging of the brains of Leborgne and Lelong. *Brain*. 2007;130(Pt 5):1432–1441.
 34. Hickok G, Poeppel D. Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*. 2004;92:67–99.
 35. Hickok G, Poeppel D. The cortical organization of speech processing. *Nat Rev Neurosci*. 2007;8:393–402.
 36. Rauschecker JP, Scott SK. Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nat Neurosci*. 2009;12:718–724.
 37. Glasser MF, Coalson TS, Robinson EC, et al. A multi-modal parcellation of human cerebral cortex. *Nature*. 2016;536:171–178.
 38. Baboyan V, Basilakos A, Yourganov G, et al. Isolating the white matter circuitry of the dorsal language stream: Connectome-symptom mapping in stroke induced aphasia. *Hum Brain Mapp*. 2021;42:5689–5702.
 39. Duffy JR. *Motor speech disorders: Substrates, differential diagnosis, and management*: Elsevier; 2013.
 40. Mailend ML, Maas E. To lump or to split? Possible subtypes of apraxia of speech. *Aphasiology*. 2020;35:592–613.
 41. Rogalsky C, Poppa T, Chen KH, et al. Speech repetition as a window on the neurobiology of auditory-motor integration for speech: A voxel-based lesion symptom mapping study. *Neuropsychologia*. 2015;71:18–27.
 42. Geschwind N. Disconnexion syndromes in animals and man. *Brain*. 1965;88:237–294, 585–644.
 43. Buchsbaum BR, Baldo J, Okada K, et al. Conduction aphasia, sensory-motor integration, and phonological short-term memory—An aggregate analysis of lesion and fMRI data. *Brain Lang*. 2011;119:119–128.
 44. Catani M, Jones DK, ffytche DH. Perisylvian language networks of the human brain. *Ann Neurol*. 2005;57:8–16.
 45. Chang EF, Raygor KP, Berger MS. Contemporary model of language organization: An overview for neurosurgeons. *J Neurosurg*. 2015;122:250–261.
 46. Zilles K, Amunts K. Cytoarchitectonic and receptorarchitectonic organization in Broca's region and surrounding cortex. *Curr Opin Behav Sci*. 2018;21:93–105.
 47. Amunts K, Mohlberg H, Bludau S, Zilles K. Julich-Brain: A 3D probabilistic atlas of the human brain's cytoarchitecture. *Science*. 2020;369:988–992.
 48. Geyer S. *The microstructural border between the motor and the cognitive domain in the human cerebral cortex*. Springer-Verlag; 2004.
 49. Eickhoff SB, Bzdok D, Laird AR, Kurth F, Fox PT. Activation likelihood estimation meta-analysis revisited. *Neuroimage*. 2012;59:2349–2361.
 50. Eichert N, Papp D, Mars RB, Watkins KE. Mapping human laryngeal motor cortex during vocalization. *Cereb Cortex*. 2020;30:6254–6269.
 51. Venezia JH, Richards VM, Hickok G. Speech-driven spectrotemporal receptive fields beyond the auditory cortex. *Hear Res*. 2021;408:108307.
 52. Venezia JH, Hickok G, Richards VM. Auditory “bubbles”: Efficient classification of the spectrotemporal modulations essential for speech intelligibility. *J Acoust Soc Am*. 2016;140:1072.
 53. Venezia JH, Thurman SM, Richards VM, Hickok G. Hierarchy of speech-driven spectrotemporal receptive fields in human auditory cortex. *Neuroimage*. 2019;186:647–666.
 54. Ozker M, Doyle W, Devinsky O, Flinker A. A cortical network processes auditory error signals during human speech production to maintain fluency. *PLoS Biol*. 2022;20:e3001493.
 55. Buchsbaum B, Hickok G, Humphries C. Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cogn Sci*. 2001;25:663–678.
 56. Brown S, Ngan E, Liotti M. A larynx area in the human motor cortex. *Cereb Cortex*. 2008;18:837–845.
 57. Yairi E, Ambrose N. Epidemiology of stuttering: 21st century advances. *J Fluency Disord*. 2013;38:66–87.
 58. Prasse JE, Kikano GE. Stuttering: An overview. *Am Fam Physician*. 2008;77:1271–1276.
 59. Guenther FH. *Neural control of speech*. MIT Press; 2016.
 60. Andrews G, Howie PM, Dozsa M, Guitar BE. Stuttering: Speech pattern characteristics under fluency-inducing conditions. *J Speech Hear Res*. 1982;25:208–216.
 61. Wan CY, Ruber T, Hohmann A, Schlaug G. The therapeutic effects of singing in neurological disorders. *Music Percept*. 2010;27:287–295.
 62. Neef NE, Primassin A, von Gudenberg AW, et al. Two cortical representations of voice control are differentially involved in speech fluency. *Brain Commun*. 2021;3:fcaa232.
 63. Neumann K, Euler HA, Kob M, et al. Assisted and unassisted recession of functional anomalies associated with dysprosody in adults who stutter. *J Fluency Disord*. 2018;55:120–134.

64. Uithol S, van Rooij I, Bekkering H, Haselager P. Hierarchies in action and motor control. *J Cogn Neurosci*. 2012;24:1077–1086.
65. Jarvis ED. Evolution of vocal learning and spoken language. *Science*. 2019;366:50–54.
66. Eichert N, Watkins KE, Mars RB, Petrides M. Morphological and functional variability in central and subcentral motor cortex of the human brain. *Brain Struct Funct*. 2021;226:263–279.
67. Simonyan K. The laryngeal motor cortex: Its organization and connectivity. *Curr Opin Neurobiol*. 2014;28:15–21.
68. Breshears JD, Molinaro AM, Chang EF. A probabilistic map of the human ventral sensorimotor cortex using electrical stimulation. *J Neurosurg*. 2015;123:340–349.
69. Belyk M, Eichert N, McGettigan C. A dual larynx motor networks hypothesis. *Philos Trans R Soc Lond Ser B, Biol Sci*. 2021;376:20200392.
70. Graziano MS, Taylor CS, Moore T, Cooke DF. The cortical control of movement revisited. *Neuron*. 2002;36:349–362.
71. Kaas JH, Gharbawie OA, Stepniewska I. Cortical networks for ethologically relevant behaviors in primates. *Am J Primatol*. 2013;75:407–414.
72. Graziano MS, Aflalo TN. Rethinking cortical organization: Moving away from discrete areas arranged in hierarchies. *Neuroscientist*. 2007;13:138–147.
73. Kaas JH. Evolution of columns, modules, and domains in the neocortex of primates. *Proc Natl Acad Sci U S A*. 2012;109(Suppl 1):10655–10660.
74. Chan PS, Cheng CH, Wu YT, et al. Neural substrates of respiratory sensory gating: A human fMRI study. *Biol Psychol*. 2022;169:108277.
75. McKay LC, Evans KC, Frackowiak RS, Corfield DR. Neural correlates of voluntary breathing in humans. *J Appl Physiol* (1985). 2003;95:1170–1178.
76. Simonyan K, Ostuni J, Ludlow CL, Horwitz B. Functional but not structural networks of the human laryngeal motor cortex show left hemispheric lateralization during syllable but not breathing production. *J Neurosci*. 2009;29:14912–14923.
77. Simonyan K, Jurgens U. Afferent cortical connections of the motor cortical larynx area in the rhesus monkey. *Neuroscience*. 2005;130:133–149.
78. Huang CS, Sirisko MA, Hiraba H, Murray GM, Sessle BJ. Organization of the primate face motor cortex as revealed by intracortical microstimulation and electrophysiological identification of afferent inputs and corticobulbar projections. *J Neurophysiol*. 1988;59:796–818.
79. Yamada Y, Yamamura K, Inoue M. Coordination of cranial motoneurons during mastication. *Respir Physiol Neurobiol*. 2005;147:177–189.
80. Ferrari PF, Gallese V, Rizzolatti G, Fogassi L. Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *Eur J Neurosci*. 2003;17:1703–1714.
81. Rizzolatti G, Camarda R, Fogassi L, Gentilucci M, Luppino G, Matelli M. Functional organization of inferior area 6 in the macaque monkey. II. Area F5 and the control of distal movements. *Exp Brain Res*. 1988;71:491–507.
82. Shepherd SV, Freiwald WA. Functional networks for social communication in the macaque monkey. *Neuron*. 2018;99:413–420.e3.
83. Coude G, Ferrari PF, Roda F, et al. Neurons controlling voluntary vocalization in the macaque ventral premotor cortex. *PLoS One*. 2011;6:e26822.
84. Hage SR, Nieder A. Single neurons in monkey prefrontal cortex encode volitional initiation of vocalizations. *Nat Commun*. 2013;4:2409.
85. Kirzinger A, Jurgens U. Cortical lesion effects and vocalization in the squirrel monkey. *Brain Res*. 1982;233:299–315.
86. Jurgens U. The neural control of vocalization in mammals: A review. *J Voice*. 2009;23:1–10.
87. Brown S, Yuan Y, Belyk M. Evolution of the speech-ready brain: The voice/jaw connection in the human motor cortex. *J Comp Neurol*. 2021;529:1018–1028.
88. Matchin W, Basilakos A, Stark BC, den Ouden DB, Fridriksson J, Hickok G. Agrammatism and paragrammatism: A cortical double dissociation revealed by lesion–symptom mapping. *Neurobiol Lang*. 2020;1:208–225.
89. Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD. Large-scale automated synthesis of human functional neuroimaging data. *Nat Methods*. 2011;8:665–670.
90. Shattuck-Hufnagel S. Prosodic frames in speech production. In: Redford MA, ed. *The handbook of speech production*. John Wiley & Sons, Inc.; 2015:419–444.
91. Shattuck-Hufnagel S, Ren A. The prosodic characteristics of non-referential co-speech gestures in a sample of academic-lecture-style speech. *Front Psychol*. 2018;9:1514.
92. LaCroix AN, Blumenstein N, Tully M, Baxter LC, Rogalsky C. Effects of prosody on the cognitive and neural resources supporting sentence comprehension: A behavioral and lesion–symptom mapping study. *Brain Lang*. 2020;203:104756.
93. Zumbansen A, Peretz I, Hebert S. Melodic intonation therapy: Back to basics for future research. *Front Neurol*. 2014;5:7.
94. Ross ED, Shayya L, Rousseau JF. Prosodic stress: Acoustic, aphasic, aprosodic and neuroanatomic interactions. *J Neurolinguistics*. 2013;26:526–551.
95. MacNeilage PF. The frame/content theory of evolution of speech production. *Behav Brain Sci*. 1998;21:499–511; discussion 511–546.
96. Fitch WT. Sequence and hierarchy in vocal rhythms and phonology. *Ann New York Acad Sci*. 2019;1453:29–46.
97. Hagoort P. On Broca, brain, and binding: A new framework. *Trends Cogn Sci*. 2005;9:416–423.
98. Herrera Ortiz C, Whittle N, Leek MR, et al. Cortical Networks for Recognition of Speech with Simultaneous Talkers. *PsyArXiv*. 2021.
99. Voon V, Droux F, Morris L, et al. Decisional impulsivity and the associative-limbic subthalamic nucleus in obsessive-compulsive disorder: Stimulation and connectivity. *Brain*. 2017;140:442–456.
100. Power JD, Plitt M, Gotts SJ, et al. Ridding fMRI data of motion-related influences: Removal of signals with distinct spatial and physical bases in multiecho data. *Proc Natl Acad Sci U S A*. 2018;115:E2105–E2114.
101. Kundu P, Brenowitz ND, Voon V, et al. Integrated strategy for improving functional connectivity mapping using multiecho fMRI. *Proc Natl Acad Sci U S A*. 2013;110:16187–16192.
102. Smith SM, Nichols TE. Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage*. 2009;44:83–98.
103. Wu J, Ngo GH, Greve D, et al. Accurate nonlinear mapping between MNI volumetric and FreeSurfer surface coordinate systems. *Hum Brain Mapp*. 2018;39:3793–3808.
104. Darwin C. *The descent of man and selection in relation to sex*. John Murray; 1871.
105. Fitch WT. *The evolution of language*. Cambridge University Press; 2010.
106. Fitch WT. Empirical approaches to the study of language evolution. *Psychon Bull Rev*. 2017;24:3–33.
107. Mithen S. *The singing Neanderthals: The origins of music, language, mind and body*. Weidenfeld & Nicolson; 2005.

108. Martins PT, Boeckx C. Attention mechanisms and the mosaic evolution of speech. *Front Psychol.* 2014;5:1463.
109. Hickok G. Computational neuroanatomy of speech production. *Nat Rev Neurosci.* 2012;13:135–145.
110. Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci.* 1999;2:1131–1136.
111. Hackett TA, Stepniewska I, Kaas JH. Prefrontal connections of the parabelt auditory cortex in macaque monkeys. *Brain Res.* 1999;817:45–58.
112. Caruso VC, Pages DS, Sommer MA, Groh JM. Similar prevalence and magnitude of auditory-evoked and visually evoked activity in the frontal eye fields: Implications for multisensory motor control. *J Neurophysiol.* 2016;115:3162–3173.
113. Thompson KG, Biscoe KL, Sato TR. Neuronal basis of covert spatial attention in the frontal eye field. *J Neurosci.* 2005;25:9479–9487.
114. Tobyne SM, Osher DE, Michalka SW, Somers DC. Sensory-biased attention networks in human lateral frontal cortex revealed by intrinsic functional connectivity. *Neuroimage.* 2017;162:362–372.
115. Assem M, Shashidhara S, Glasser MF, Duncan J. Precise topology of adjacent domain-general and sensory-biased regions in the human brain. *Cereb Cortex.* 2022;32:2521–2537.
116. Brown CH, Schessler T, Moody D, Stebbins W. Vertical and horizontal sound localization in primates. *J Acoust Soc Am.* 1982;72:1804–1811.
117. Cabrera D, Morimoto M. Influence of fundamental frequency and source elevation on the vertical localization of complex tones and complex tone pairs. *J Acoust Soc Am.* 2007;122:478–488.
118. Parise CV, Knorre K, Ernst MO. Natural auditory scene statistics shapes human spatial hearing. *Proc Natl Acad Sci U S A.* 2014;111:6104–6108.
119. Chiou R, Rich AN. Cross-modality correspondence between pitch and spatial location modulates attentional orienting. *Perception.* 2012;41:339–353.
120. Roffler SK, Butler RA. Localization of tonal stimuli in the vertical plane. *J Acoust Soc Am.* 1968;43:1260–1266.
121. Walker P, Bremner JG, Lunghi M, Dolscheid S B DB, Simion F. Newborns are sensitive to the correspondence between auditory pitch and visuospatial elevation. *Develop Psychobiol.* 2018;60:216–223.
122. Korzeniowska AT, Root-Gutteridge H, Simner J, Reby D. Audio-visual crossmodal correspondences in domestic dogs (*Canis familiaris*). *Biol Lett.* 2019;15:20190564.
123. Fitch WT. Skull dimensions in relation to body size in non-human mammals: The causal basis for acoustic allometry. *Zoology.* 2000;103:40–58.
124. Pisanski K, Cartei V, McGettigan C, Raine J, Reby D. Voice modulation: A window into the origins of human vocal control? *Trends Cogn Sci.* 2016;20:304–318.
125. Kayser C, Petkov CI, Lippert M, Logothetis NK. Mechanisms for allocating auditory attention: An auditory saliency map. *Curr Biol.* 2005;15:1943–1947.
126. Pisanski K, Isenstein SGE, Montano KJ, O'Connor JJM, Feinberg DR. Low is large: Spatial location and pitch interact in voice-based body size estimation. *Atten Percept Psychophys.* 2017;79:1239–1251.
127. Kirchner H, Barbeau EJ, Thorpe SJ, Regis J, Liegeois-Chauvel C. Ultra-rapid sensory responses in the human frontal eye field region. *J Neurosci.* 2009;29:7599–7606.
128. Romanski LM, Goldman-Rakic PS. An auditory domain in primate prefrontal cortex. *Nat Neurosci.* 2002;5:15–16.
129. Graziano MS, Reiss LA, Gross CG. A neuronal representation of the location of nearby sounds. *Nature.* 1999;397:428–430.
130. Hage SR. Auditory and audio-vocal responses of single neurons in the monkey ventral premotor cortex. *Hear Res.* 2018;366:82–89.
131. Loh KK, Petrides M, Hopkins WD, Procyk E, Amiez C. Cognitive control of vocalizations in the primate ventrolateral-dorsomedial frontal (VLF-DMF) brain network. *Neurosci Biobehav Rev.* 2017;82:32–44.
132. Brenowitz EA. Evolution of the vocal control system in the avian brain. *Seminars Neurosci.* 1991;3:399–407.
133. Jancke L, Loose R, Lutz K, Specht K, Shah NJ. Cortical activations during paced finger-tapping applying visual and auditory pacing stimuli. *Brain Res Cogn Brain Res.* 2000;10:51–66.
134. Pollok B, Krause V, Butz M, Schnitzler A. Modality specific functional interaction in sensorimotor synchronization. *Hum Brain Mapp.* 2009;30:1783–1790.
135. Chen JL, Zatorre RJ, Penhune VB. Interactions between auditory and dorsal premotor cortex during synchronization to musical rhythms. *Neuroimage.* 2006;32:1771–1781.
136. Baumann S, Koeneke S, Schmidt CF, Meyer M, Lutz K, Jancke L. A network for audio-motor coordination in skilled pianists and non-musicians. *Brain Res.* 2007;1161:65–78.
137. Archakov D, DeWitt I, Kusmierek P, et al. Auditory representation of learned sound sequences in motor regions of the macaque brain. *Proc Natl Acad Sci U S A.* 2020;117:15242–15252.
138. Vernet M, Quentin R, Chanes L, Mitsumasu A, Valero-Cabre A. Frontal eye field, where art thou? Anatomy, function, and non-invasive manipulation of frontal regions involved in eye movements and associated cognitive operations. *Front Integr Neurosci.* 2014;8:66.
139. Tu TA, Keating EG. Electrical stimulation of the frontal eye field in a monkey produces combined eye and head movements. *J Neurophysiol.* 2000;84:1103–1106.
140. Petit L, Beauchamp MS. Neural basis of visually guided head movements studied with fMRI. *J Neurophysiol.* 2003;89:2516–2527.
141. Posner MI. Orienting of attention: Then and now. *Q J Exp Psychol (Hove).* 2016;69:1864–1875.
142. Posner MI, Niell CM. Illuminating the neural circuits underlying orienting of attention. *Vision (Basel).* 2019;3:4.
143. Thompson WF, Russo FA. Facing the music. *Psychol Sci.* 2007;18:756–757.
144. Knight EJ, Austin SF. The effect of head flexion/extension on acoustic measures of singing voice quality. *J Voice.* 2020;34:964.e911–964.e921.
145. Munhall KG, Jones JA, Callan DE, Kuratate T, Vatikiotis-Bateson E. Visual prosody and speech intelligibility: Head movement improves auditory speech perception. *Psychol Sci.* 2004;15:133–137.
146. Burnham D, Vatikiotis-Bateson E, Barbosa AV, et al. Seeing lexical tone: Head and face motion in production and perception of Cantonese lexical tones. *Speech Commun.* 2022;141:40–55.
147. Loehr DP. Temporal, structural, and pragmatic synchrony between intonation and gesture. *Lab Phonol.* 2012;3:71–89.
148. Ghazanfar AA, Takahashi DY, Mathur N, Fitch WT. Cineradiography of monkey lip-smacking reveals putative precursors of speech dynamics. *Curr Biol.* 2012;22:1176–1182.
149. Ghazanfar AA, Takahashi DY. Facial expressions and the evolution of the speech rhythm. *J Cogn Neurosci.* 2014;26:1196–1207.
150. Kohler E, Keysers C, Umiltà MA, Fogassi L, Gallese V, Rizzolatti G. Hearing sounds, understanding actions: Action representation in mirror neurons. *Science.* 2002;297:846–848.
151. Jurgens U. Neural pathways underlying vocal control. *Neurosci Biobehav Rev.* 2002;26:235–258.

152. Hickok G. A cortical circuit for voluntary laryngeal control: Implications for the evolution language. *Psychon Bull Rev.* 2017;24:56–63.
153. Parrell B, Ramanarayanan V, Nagarajan S, Houde J. The FACTS model of speech motor control: Fusing state estimation and task-based control. *PLoS Comput Biol.* 2019;15:e1007321.
154. Thorson JC. The role of prosody in early word learning. In: Prieto P, Esteve-Gilbert N, eds. *The development of prosody in first language acquisition*. John Benjamins Publishing Company; 2018:59–78.
155. Nieder A, Mooney R. The neurobiology of innate, volitional and learned vocalizations in mammals and birds. *Philos Trans R Soc Lond Ser B Biol Sci.* 2020;375:20190054.
156. Patel AD. Vocal learning as a preadaptation for the evolution of human beat perception and synchronization. *Philos Trans R Soc Lond Ser B Biol Sci.* 2021;376:20200326.
157. Nathani S, Ertmer DJ, Stark RE. Assessing vocal development in infants and toddlers. *Clin Linguist Phon.* 2006;20:351–369.
158. Cho-Hisamoto Y, Kojima K, Brown EC, Matsuzaki N, Asano E. Cooing- and babbling-related gamma-oscillations during infancy: Intracranial recording. *Epilepsy Behav.* 2012;23:494–496.
159. Bergman TJ. Speech-like vocalized lip-smacking in geladas. *Curr Biol.* 2013;23:R268–R269.
160. Watson SK, Townsend SW, Schel AM, et al. Vocal learning in the functionally referential food grunts of chimpanzees. *Curr Biol.* 2015;25:495–499.
161. Lameira AR, Hardus ME, Mielke A, Wich SA, Shumaker RW. Vocal fold control beyond the species-specific repertoire in an orang-utan. *Sci Rep.* 2016;6:30315.
162. Ponsot E, Burred JJ, Belin P, Aucouturier JJ. Cracking the social code of speech prosody using reverse correlation. *Proc Natl Acad Sci.* 2018;115:3972–3977.
163. Charlton BD, Reby D. The evolution of acoustic size exaggeration in terrestrial mammals. *Nat Commun.* 2016;7:12739.
164. Donald M. Key cognitive preconditions for the evolution of language. *Psychon Bull Rev.* 2017;24:204–208.
165. Molesti S, Meguerditchian A, Bourjade M. Gestural communication in olive baboons (*Papio anubis*): Repertoire and intentionality. *Anim Cogn.* 2019;23:19–40.
166. Arbib MA. From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behav Brain Sci.* 2005;28:105–124; discussion 125–167.
167. Corballis MC. Mirror neurons and the evolution of language. *Brain Lang.* 2010;112:25–35.
168. de Boer B. Evolution of speech and evolution of language. *Psychon Bull Rev.* 2017;24:158–162.
169. Goldin-Meadow S. What the hands can tell us about language emergence. *Psychon Bull Rev.* 2017;24:213–218.
170. Kendon A. Reflections on the “gesture-first” hypothesis of language origins. *Psychon Bull Rev.* 2017;24:163–170.
171. Hubscher I, Prieto P. Gestural and prosodic development act as sister systems and jointly pave the way for children’s socio-pragmatic development. *Front Psychol.* 2019;10:1259.
172. Ross ED, Monnot M. Neurology of affective prosody and its functional–anatomic organization in right hemisphere. *Brain Lang.* 2008;104:51–74.
173. Ukaegbe OC, Holt BE, Keator LM, Brownell H, Blake ML, Lundgren K. Aprosodia following focal brain damage: What’s right and what’s left? *Am J Speech-Lang Pathol.* 2022;31:2313–2328.
174. Wright AE, Davis C, Gomez Y, et al. Acute ischemic lesions associated with impairments in expression and recognition of affective prosody. *Perspect ASHA Spec Interest Groups.* 2016;1:82–95.
175. Belkhir JR, Fitch WT, Garcea FE, et al. Direct electrical stimulation evidence for a dorsal motor area with control of the larynx. *Brain Stimul.* 2021;14:110–112.
176. Hamilton LS, Edwards E, Chang EF. A spatial map of onset and sustained responses to speech in the human superior temporal gyrus. *Curr Biol.* 2018;28:1860–1871.e4.
177. Berezutskaya J, Baratin C, Freudenburg ZV, Ramsey NF. High-density intracranial recordings reveal a distinct site in anterior dorsal precentral cortex that tracks perceived speech. *Hum Brain Mapp.* 2020;41:4587–4609.
178. Ozdemir E, Norton A, Schlaug G. Shared and distinct neural correlates of singing and speaking. *Neuroimage.* 2006;33:628–635.
179. Royal I, Lidji P, Theoret H, Russo FA, Peretz I. Excitability of the motor system: A transcranial magnetic stimulation study on singing and speaking. *Neuropsychologia.* 2015;75:525–532.
180. Poeppel D. The analysis of speech in different temporal integration windows: Cerebral lateralization as “asymmetric sampling in time”. *Speech Commun.* 2003;41:245–255.
181. Zatorre RJ, Belin P, Penhune VB. Structure and function of auditory cortex: Music and speech. *Trends Cogn Sci.* 2002;6:37–46.