

UCSF

UC San Francisco Previously Published Works

Title

The Control of Vocal Pitch in Human Laryngeal Motor Cortex

Permalink

<https://escholarship.org/uc/item/6946p7w9>

Journal

Cell, 174(1)

ISSN

0092-8674

Authors

Dichter, Benjamin K
Breshears, Jonathan D
Leonard, Matthew K
[et al.](#)

Publication Date

2018-06-01

DOI

10.1016/j.cell.2018.05.016

Peer reviewed



HHS Public Access

Author manuscript

Cell. Author manuscript; available in PMC 2019 June 28.

Published in final edited form as:

Cell. 2018 June 28; 174(1): 21–31.e9. doi:10.1016/j.cell.2018.05.016.

The control of vocal pitch in human laryngeal motor cortex

Benjamin K. Dichter^{1,2,3}, Jonathan D. Breshears^{1,2}, Matthew K. Leonard^{1,2}, and Edward F. Chang^{1,2,3}

¹Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, California 94158, USA

²Department of Neurological Surgery, University of California, San Francisco, San Francisco, California 94143, USA

³UC Berkeley and UCSF Joint Program in Bioengineering, Berkeley, California 94720, USA

Summary

In speech, the highly flexible modulation of vocal pitch creates intonation patterns that speakers use to convey linguistic meaning. This human ability is unique among primates. Here, we used high-density cortical recordings directly from the human brain to determine the encoding of vocal pitch during natural speech. We found neural populations in bilateral dorsal laryngeal motor cortex (dLMC) that selectively encoded produced pitch, but not non-laryngeal articulatory movements. This neural population controlled short pitch accents to express prosodic emphasis on a word in a sentence. Other larynx cortical representations controlling voicing and longer pitch phrase contours were found at separate sites. dLMC sites also encoded vocal pitch during a non-speech singing task. Finally, direct focal stimulation of dLMC evoked laryngeal movements and involuntary vocalization, confirming its causal role in feedforward control. Together, these results reveal the neural basis for the voluntary control of vocal pitch in human speech.

Graphical Abstract

Corresponding Author and Lead Contact, Edward.Chang@ucsf.edu.

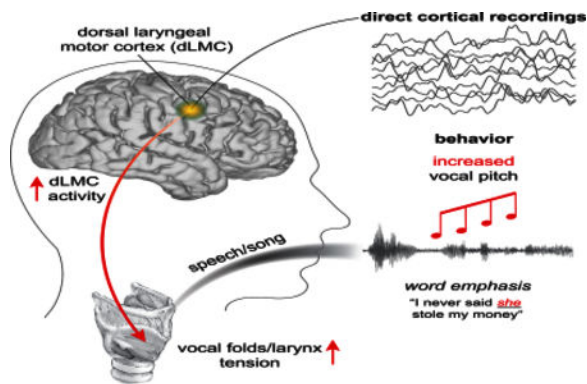
Author Contributions

B.K.D. contributed experimental design, data collection, data analysis, and writing the manuscript; J.D.B contributed experimental design, data collection, and writing the manuscript; M.K.L. contributed data visualization and writing the manuscript; E.F.C. contributed to experimental design, data collection, writing the manuscript, and project supervision.

Declaration of Interests

The authors claim no competing interest.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Introduction

The precise control of the larynx is central to the human ability to speak. In English, for example, deliberately controlled changes of vocal pitch are used to convey critical elements of prosody, including syllable stress, word emphasis, phrase segmentation, modality (e.g. question vs. statement), and mood (Ladd, 2008).

In speech, the two dominant functions of the larynx are to generate voicing and modulate pitch. Voicing is created by bringing the vocal folds into close proximity, so that they vibrate when air is passed through. In contrast, the pitch of the voice is modulated primarily by fine changes in the tension of the vocal folds. Greater tension in the vocal folds causes them to vibrate at a higher frequency during voicing, producing a higher pitch sound (Hull, 2013; Titze et al., 1989). The fine control of pitch that gives rise to intonational patterns during speech (Collier, 1975) and melodies in singing (Roubeau et al. 1997) is primarily mediated by flexing the cricothyroid muscle (Figure 1A), which tilts the thyroid cartilage with respect to the cricoid cartilage, stretching the vocal folds.

The ability to voluntarily and flexibly control pitch patterns, in the context of vocal learning, is unique to humans among primates (Kirzinger and Jürgens, 1982; Simonyan, 2013; Belyk and Brown, 2017). While it was previously thought that this ability was due to anatomical differences in the larynx (Lieberman et al., 1969), recent evidence suggests that evolutionary changes in neural control of vocalizations may play a key role in enabling speech behavior (Fitch et al., 2016; Jürgens, 2002; Simonyan, 2013). It has been speculated that the cortical control of complex laryngeal function was a key factor in enabling flexible expression of prosody, and thereby was a driving force behind the rapid development of spoken language in humans (Belyk and Brown, 2017; Brown et al., 2008; Hickok, 2016; Pisanski et al., 2016).

Recent studies have identified two distinct regions in the human sensorimotor cortex that are correlated with laryngeal movements (Bouchard et al. 2013). The ventral laryngeal motor cortex (vLMC) is at the bottom of the sensorimotor cortex homunculus (Foerster, 1936; Penfield and Boldrey, 1937), and has been previously described as a homologue of LMC in other primate species (Hast and Milojkvic, 1966; Jürgens, 2009; Simonyan and Jürgens, 2002; Ludlow, 2005). A completely separate dorsal premotor region (dLMC) has been more

recently identified between the cortical representation of the lips and the hand (Belyk and Brown, 2015; Brown et al., 2008; Olthoff et al., 2008; Rodel et al., 2004; Simonyan and Horwitz, 2011). The existence of two distinct larynx cortical representations in the sensorimotor cortex is controversial (Belyk and Brown, 2017), in part because it is unknown how and whether each region contributes to distinct roles in larynx control.

The larynx motor cortex is part of a larger vocal tract sensorimotor cortex that is critical for fluent speech, as injury to these areas can cause dysarthria and apraxia of speech (Patira et al., 2017; Wilson et al., 2015). In contrast, ablating this area in non-human primates has no apparent effect on vocalization behavior (Kirzinger and Jürgens, 1982). Despite its role in speech production, fundamental questions exist about what information is represented there, specifically with regard to vocal pitch as one of the core laryngeal functions in speech production. Candidate representations include the control of larynx movements, their acoustic sensory goals, or even a high-order linguistic-level encoding of prosody.

To understand how cortical activity relates to pitch control, we sought to determine: 1) whether there is separable encoding for functionally-distinct pitch components such as accents and phrase, as well as voicing, 2) whether the same pitch control mechanisms are engaged during non-speech vocalizations, like singing a melody, and 3) whether dLMC activity reflects the causal, feed-forward, and proportional control of laryngeal muscles. While previous human imaging studies have focused on where laryngeal representations are localized in the cortex, our goal here was to determine the encoding of vocal pitch using methods with high spatial resolution but also simultaneous high temporal fidelity to resolve the rapid and flexible dynamics of pitch changes in natural speech. Thus, we used high-density intracranial recordings and stimulation of the lateral surface of the human brain in participants who were undergoing epilepsy surgery. These high-resolution recordings allowed us to identify the specific functional roles of both LMC regions during naturalistic vocal production tasks.

Results

Pitch encoding in the dLMC during natural speech intonation

To understand how speakers control the pitch of their voices, we first designed a lexical emphasis task during which participants stressed specific words in a sentence. Twelve participants spoke the sentence, “I never said she stole my money.” On each trial, they were cued to change the meaning of the sentence by emphasizing a specific word (Rooth, 1992) (Figure 1B). For instance, “I never said *she* stole my money” implies that someone else had stolen the money. On each trial, the written forms of the sentences were presented to participants on a computer screen with the target word of emphasis underlined and italicized, and an example audio sentence was played through speakers. In an additional to emphasizing each word, on some trials participants were instructed to say the sentence as a question. This task naturally elicits prosodic differences between conditions, while keeping the lexical and syllabic content of each sentence the same.

We used an autocorrelation method to extract the pitch contour (fundamental frequency, f_0) from the produced acoustic waveform (Boersma, 1993). On every trial, the pitch contour

contained a transient increase in pitch at the time of the emphasized word (Figure 1B, green lines). While participants performed the task, we recorded neural activity from high density electrocorticography (ECoG) arrays (256 electrodes) that broadly covered speech-processing areas across the lateral cortical surface (Figure S1). Note that electrode coverage was only unilateral for clinical reasons (left N=6, right N=6), so we could not directly compare laterality of LMC function in a single participant. We computed the analytic amplitude of the cortical activity in the high-gamma range (HG; 70–150 Hz), which has been found to correlate with multi-unit neuronal firing rates (Ray and Maunsell, 2011), and has been shown to reliably track neural activity associated with speech articulation and other movements (Bouchard et al., 2013; Conant et al., 2018; Crone et al., 1998).

During this lexical emphasis task, we found electrodes with increased neural activity that was clearly time-locked to the production of the emphasized word (Figure 1B, single trial raster plots). This neural activation started at the onset or slightly in advance of the emphasized word. We next quantified the relationship between vocal pitch and neural activity for every electrode across participants. To control for the encoding of supralaryngeal articulators (e.g. lips, tongue, jaw, etc.) (Bouchard et al., 2013), we first used dynamic time warping on the acoustics to temporally align the syllabic sequence across trials, then subtracted the mean activation pattern across trials. To understand the encoding of pitch in the context of word emphasis, we also controlled for natural declination of pitch (Ladd, 1984), which causes a correlation between pitch and proximity to the start of the sentence. A trial-wise shuffle test was used, and electrodes that correlate more strongly with pitch than would be expected from declination alone were considered significantly encoding intonation (See methods for details).

After removing these potential confounds, we found that across participants, electrodes that were significantly correlated with the produced pitch were specifically localized to a region of the precentral gyrus, the dorsal laryngeal motor cortex (dLMC; $p < 0.001$ using a shuffle test; Figure 1C). These electrodes were found bilaterally on both the right and left hemispheres. On the right hemisphere, the electrodes were tightly clustered in the dLMC across participants. On the left hemisphere, they appeared in the homologous location, but we also observed some pitch-encoding electrodes at other locations in the ventral sensorimotor cortex (vSMC). Similar results were observed when also using partial correlation to remove the effect of intensity (amplitude) (Stevens, 1935) (Figure S1). All pitch-encoding electrodes showed a positive monotonic relationship between neural activity and pitch (Figure 1D). That is, increases in high gamma activity were followed by linear and proportional increases in produced pitch. For comparison, we also observed evidence for encoding of the auditory feedback of vocal pitch in electrodes over the bilateral non-primary auditory cortices (superior temporal gyrus (STG)). In contrast to dLMC, which only had positive correlations, STG electrodes appeared to be directionally tuned to either increases or decreases in self-produced pitch (Figure 1E).

We next wanted to confirm that the results for this lexical emphasis task would generalize to other natural speaking conditions. Participants were asked to speak aloud 50 whole sentences selected from the MOCHA-TIMIT corpus (Wrench, 1999), reading from a computerized prompt. In this control condition, participants were asked to speak naturally,

without any specific instructions for intonation. We found a close correspondence between pitch-encoding electrodes in the dLMC for MOCHA-TIMIT sentences and the lexical emphasis task (Figure S1). Furthermore, we found that pitch-encoding electrodes in the lexical emphasis task also predicted the neural activity at many of the same electrodes in the MOCHA reading task (Pearson $r = 0.33$; p -value $< 1e-4$) (Figure S1).

dLMC has both motor and auditory sensory representations

Since neural activations were broadly concurrent with intonation, we next considered whether they represented sensory feedback versus motor feed-forward commands. It has been previously demonstrated that areas of the speech motor cortex can be active during listening (Cheung et al., 2016; Wilson et al., 2004), more active during auditory pitch discrimination (Sammler et al., 2015) and when planning to repeat a melody (Nishida et al., 2017), and disruption of this area decreases pitch discrimination performance (Sammler et al., 2015). To evaluate whether the pitch-correlated activity observed in our speech tasks might reflect only the auditory feedback response to one's own speech, we recorded individual participants' productions and played back their own speech passively through audio speakers (N=4 participants). We found that dLMC electrodes that were correlated with pitch during speaking also had auditory responses that were correlated with pitch during listening, although with distinct response latencies. The cross-correlation response of the example electrode (Figure 1F) showed a positive correlation with pitch in the listening condition only after the pitch increase, peaking at about 0.3 seconds, 0.1 seconds after the peak of the cross-correlation during speaking. Across all dLMC electrodes that were correlated in both speaking and listening conditions, the cross-correlation value surpassed 0.1 later in the listening condition than in the speaking condition by 0.1 seconds on average (p -value=0.019; 1-sided student t-test).

Next, we examined the timing of dLMC activity at the onset of sentences for speaking and listening conditions. We compared the neural activity around the beginning of the word "I", across all sentence styles. Activation was strongly time-locked to sentence onset, as illustrated by the example electrode (Figure 1G), with the activation in the speaking condition occurring before the playback condition. For each electrode in the dLMC, we determined when the mean activation crossed a 1 s.d. threshold in the speaking and playback tasks. We found that most electrodes activated before acoustic onset when speaking, with a mean lead of 0.09 s, and that when listening, all electrodes responded after acoustic onset, with a mean lag of 0.39 s. All electrodes in dLMC that responded to both speaking and listening responded later in the listening task than in the speaking task, and the difference between the speaking and listening activation times was statistically significant (paired t-test, p -value < 0.001). By directly comparing these response latencies at single electrodes, we found that dLMC has sensorimotor functions with both auditory and motor representations for vocal pitch, which has not been observed for other parts of the ventral sensorimotor cortex (vSMC) (Cheung et al., 2016).

dLMC and vLMC encoding of pitch components: accent, phrase, and voicing

In natural speech, vocal pitch is composed of multiple elements, each with different timescales, and potentially with different encoding mechanisms. To understand the specific

sub-processes involved in pitch control, we applied a model-based approach to estimate these distinct components of the pitch contour. We adapted a well-known mathematical formalization (Fujisaki, 2004), called the Fujisaki model, to explain the neural activity on each electrode in the vSMC. For each sentence, the components in the model include a fast “accent” component (emphasized words or syllables), and a slow “phrase” component (the declination (Ladd, 1984) in pitch over the course of a phrase). The model is motivated by the physiological mechanisms of pitch control in the larynx, and is capable of parsimoniously modeling pitch contours across many languages (Fujisaki, 2004). We hypothesized that these theoretically distinct components are controlled independently in the brain.

Modeling the pitch contours of each sentence as phrase and accent components, along with whether the segments were voiced or unvoiced, allowed us to reconstruct the produced pitch contours nearly perfectly ($R^2=0.96$, Figure 2A). At individual electrodes, high-gamma activity was correlated with these pitch components in a temporally-specific fashion (Figure 2B-D). Crucially, we found a clear and striking dissociation between electrodes that encoded accent, phrase, and voicing (Figure 2E-F). 66% of accent-tuned electrodes were not tuned to voicing, and 71% of voicing electrodes were not tuned to accent, suggesting that these components have separable control representations in vSMC (See Figure S2 for details).

Given the results shown in Figure 1, we hypothesized that pitch encoding in the dLMC more strongly reflects the pitch accent component, consistent with the emphasized word in each sentence. We confirmed that a subset of dLMC electrodes were most strongly associated with pitch accent (Figure 2G). In contrast, phrase-encoding electrodes were found in the vLMC and dLMC. Finally, voicing was localized to a distinct subset of dLMC and vLMC (Bouchard et al., 2013) electrodes. Together, these results demonstrate a functional-anatomical distinction between the dorsal and ventral LMCs, as well as evidence for independent and heterogeneous encoding for different pitch components within the dLMC.

dLMC pitch encoding for singing

We next asked whether the encoding of vocal pitch was specific to the linguistic context (Mayer et al., 2002), or similar during speaking and singing, a form of non-speech vocal production. Participants performed a singing task in which they listened to and then repeated pitch patterns alternating between sol-mi-do-mi-sol (high-middle-low-middle-high) and do-mi-sol-mi-do (low-middle-high-middle-low) on a vowel. Figure 3A shows examples of the two pitch melodies sung by one of the participants (Figure S3 shows the performance of all participants). To remove any effects of the sequential order of the produced pitches, the two melodies were interleaved so that the high and low notes occur in the same sequential order, both occurring third in the sequence 50% of the time and fifth in the sequence the other 50% of the time. The first note in each melody was excluded from analysis, so that all analyzed notes were preceded by the same (middle) note. Importantly, this task was specifically designed to allow us to examine pitch control while experimentally avoiding some of the potential confounds that were statistically controlled for in the natural speech experiments described earlier. That is, this singing task did not have pitch declination (Zatorre and Baum, 2012), correlations between pitch and intensity (Figure S3), or correlation between pitch and articulatory gesture, each of which are prevalent in natural speech.

In order to sing the correct pitch at the acoustic onset of each note, a singer must tense the laryngeal muscles, creating the necessary tension in the vocal folds before exhalation (Figure 1A). We did not observe this behavior in the speaking experiment, where the starting pitch for the sentence was approximately the same even when “T” is emphasized (Figure 1). However, this behavior was observed in the singing task (Figure 3A). We examined neural activity time-locked to the onset of each note (Figure 3B) and found electrodes in the anterior part of dLMC that exhibited pitch-specific activity immediately preceding the acoustic onset of the vocalization (yellow region; Figure 3C). In this brief moment before acoustic onset, we observed neural control of the larynx without participants hearing their own voice. Approximately 100–300ms after acoustic onset, electrodes in the posterior part of the dLMC were correlated with pitch (blue region; Figure 3C). Both subgroups of electrodes, those that were tuned to pitch before acoustic onset, and those that are tuned to pitch during vocalization, were also correlated with pitch during the speaking task (Figure 3C). Pitch representation was weak in the vLMC both before and during vocalization.

To quantify the similarity of pitch representation during singing and speaking, we compared the continuous correlation between electrode activity and pitch in the two conditions (Figure 3D). Across all dLMC electrodes, there was a strong correlation between encoding strength for pitch in the singing and speaking tasks (Pearson $r = 0.33$, p -value < 0.01) (Figure 3D). This demonstrates that dLMC activity reflects a task-independent representation of vocal pitch that is not specific to speech or singing. This demonstrates that dLMC activity reflects a task-independent representation of vocal pitch that is not specific to speech or singing, and may therefore reflect feed-forward control of specific laryngeal movements.

Direct electrical stimulation of dLMC evokes larynx movement and vocalization

We have demonstrated that neural activity in dLMC reflects the detailed and temporally specific features of produced pitch during speaking and singing. To definitively demonstrate that this activity reflects feed-forward control of laryngeal muscles, we used direct focal (bipolar) electrical stimulation during intraoperative clinical brain mapping. In two separate experiments, we examined whether there is a causal link between dLMC activity and laryngeal muscle activation. This approach helps establish that dLMC representations are not purely somatosensory feedback (Guenther, 2006), an efference copy signal (Niziolek et al., 2013), or an auditory response to the acoustics of one’s own voice (Behroozmand et al., 2015; Brown et al., 2008; Chang et al., 2013; Cheung et al., 2016; Wilson et al., 2004).

In the first stimulation experiment, participants undergoing neurosurgical procedures with general anesthesia were intubated with a specialized endotracheal tube with electromyographic (EMG) non-penetrating wire electrodes (Eisele, 1996; Rea and Khan, 1998). These electrodes contacted the left and right vocal folds and were designed to record laryngeal muscle activations. In 18 participants (5 left), we stimulated cortical sites throughout the sensorimotor cortex (Tate et al., 2013) while recording laryngeal EMG. We found sites that elicited a laryngeal EMG response bilaterally in the dLMC, but also sometimes in the vLMC. The highest concentration was in the dLMC, the same cortical region that correlated with vocal pitch during speech and singing (Figure 4A). The dLMC

was typically found between areas where stimulation evoked EMG-detected movements from the hand/arm (dorsal), and mouth (ventral) (Figure 4D).

To understand whether there is a causal relationship between the amount of cortical activity and the amount of laryngeal muscle activation, we varied the cortical stimulation amplitude, and found that it caused a proportional increase in the laryngeal EMG response (Pearson $r=0.85$, $p\text{-value} < 1e\text{-}52$) (Figure 4B) with a latency of 11–19 ms (Figure 4G). This demonstrates a monotonic relationship between dLMC neural activity and the magnitude of laryngeal muscle activation (Figure 4C). One example participant (red) received 11 cortical stimulations at mid-range (60 V), which elicited a distribution of laryngeal responses in between the lowest and highest stimulation magnitude. These findings of proportional responses to graded stimulation are concordant with the monotonic relationship between cortical high gamma activity and vocal pitch, which is determined by tension of the cricothyroid muscle.

In the second stimulation experiment, we asked whether stimulating the pitch-encoding region of dLMC would actually cause vocalizations in awake participants. In this experiment, stimulation was applied throughout the sensorimotor cortex in 82 neurosurgical patients undergoing awake craniotomies in which the left hemisphere cortical surface was exposed. While we could not assess the right hemisphere (awake mapping is not routinely done on the non-dominant right side), we were still interested in understanding what effects could be ascribed to dLMC stimulation given that we did find evidence of voicing encoding bilaterally. In a subset of 20 participants, we observed that stimulation of dLMC evoked audible involuntary vocalizations (Breshears et al., 2015).

We found that the evoked vocalizations were all voiced as demonstrated by energy at the fundamental frequency and voice-related harmonics (Figure 4F). These nonvolitional, stimulation-evoked vocalizations were not meaningful or communicative speech sounds, but sounded typically like a prolonged “aaah” that varied in vocal register, including vocal fry (9 participants, example: Fig. S4A), modal register (10 participants example: Figure 4F), and falsetto register (1 participant, example: Figure S4B), and lasted 0.5–2.9 seconds (mean: 1.1 seconds).

In early descriptions of evoked vocalizations by Penfield (Penfield and Roberts, 1959), similar responses were interpreted at positions spread throughout the ventral sensorimotor cortex. A distinct dorsal representation of the larynx was never depicted in historical descriptions of the homunculus. Using the precision of an intraoperative stereotactic navigation system and EMG monitoring, however, we found that these responses were well-localized to the dLMC (Figure 4E). Concordant with the EMG results, we found that stimulation at other sensorimotor cortex locations instead evoked contralateral pulling of the face, deviation of tongue, and jaw movements, or arm movements (Breshears et al., 2015). We did not observe any instances where stimulating the vLMC elicited vocalization, though previous studies have shown speech arrest from stimulating in this location (Chang et al., 2016).

These results provide definitive evidence that dLMC neural activity reflects the feedforward encoding of motor commands in the larynx, though they also suggest that the representation might be more complex than control of a single muscle or even the larynx alone. The vocalization response requires adduction of the vocal folds, but also involves precise coordination with respiratory processes in the lungs and diaphragm. The relative timing of the responses is in accordance with speech, where the larynx moves into closed position before exhaling can produce voicing and pitch (Figure 1A).

Discussion

We combined high-resolution cortical physiology and stimulation methods to demonstrate that neural signals in human dLMC encode motor commands that allow for the flexible, feed-forward control of vocal pitch. The key novel findings are as follows: 1) vocal pitch is encoded by neural activity in the bilateral dLMC, 2) dLMC electrodes encode both motor and auditory pitch-related responses, 3) accent, phrase, and voicing functions of the larynx can be separately encoded, 4) dLMC pitch encoding is similar for speech and non-speech singing, and 5) electrical stimulation of the dLMC evokes larynx movement and involuntary vocalization. These results demonstrate how prosody, a major aspect of speech production, is enabled by highly specialized sensorimotor neural control of laryngeal function in the human brain.

Relatively little was known about how the dLMC encodes larynx function, as the anatomy and physiology of the corticobulbar system is understudied in neuroscience in general and its functions in the context of speaking can only be studied directly in humans. The spatial and temporal resolution of high-density intracranial recordings facilitated our ability to resolve cortical activity at the relevant time scale of fast and transient intonation changes in speech, while also addressing differences in specific local encoding within the dLMC and vLMC regions. This allowed us to dissect various aspects of vocal pitch that have not been previously investigated.

We were specifically interested in comparing potential models of dLMC representation. For example, a high-order linguistic binary representation might code for stress (or no stress) at particular words in our task. This interpretation was ruled out because the same neural encoding was observed during the non-speech singing task, and because variability in produced pitch was directly proportional to cortical activity. Alternatively, the dLMC could be representing an auditory representation where specific electrodes encode different absolute pitch values (e.g. as in spectral receptive fields) or directional pitch changes. Indeed, we confirmed evidence for that kind of encoding in the auditory cortex STG responses here and in previous work (Tang et al., 2017), but not in the dLMC. Instead, the consistently positive monotonic relationship between pitch values and neural activity at dLMC sites suggests a motor-based model for control of pitch, perhaps reflecting muscle tension. This interpretation is consistent with previous imaging studies that localized dLMC during volitional, non-vocal larynx movements (Brown et al., 2008; Loucks et al. 2007).

However, the encoding of larynx commands does not appear to be general, but rather is quite specialized for specific modes of vocal control. By modeling distinct components of the

pitch contour in speech, the high-density recordings permitted us to functionally dissociate accent, phrase, and voicing at different discrete sites. This demonstrates how multiple dimensions of vocal pitch can be independently controlled by the cortex. This may have direct implications for temporally-precise control of pitch that involves independent processes over short (accent) and long (phrase) timescales. Voicing and pitch activate different laryngeal muscles and actions: voicing is mediated primarily by adduction of the vocal folds, and pitch is mediated by the stretching of the vocal folds. However, it was previously unknown whether and how cortical laryngeal control signals differentiated these important functions (Belyk and Brown, 2017). Consistent with our previous work (Bouchard et al., 2013), voicing was encoded by both dLMC and vLMC regions. Here, we found that a subset of electrodes within dLMC was selective for vocal pitch control, and not for other articulatory or laryngeal features, demonstrating a distinct circuit for pitch.

While the high resolution of intracranial recordings in humans here has elucidated several novel aspects of LMC function in speech, fundamental discoveries from related behaviors in animal models are likely to provide critical details to these processes in a comparative context. For example, previous research on the corticospinal tract anatomy has suggested distinct topographical subdivisions of motor cortex that feature direct versus indirect control of arm movements (Rathelot and Strick, 2009). Direct connectivity has been suggested to be a phylogenetic development in support of skilled and complex movements — such mechanisms may underlie the highly flexible control of pitch used by humans but not other primates during vocalizations. In contrast, voicing is a fundamental element of vocalizations across many species, and therefore may represent a more conserved, and integrative behavior that is coupled with respiratory function.

There is previous evidence to suggest direct LMC anatomical connectivity to laryngeal motoneurons in the nucleus ambiguus in humans (Kuypers, 1958; Kirzinger and Jürgens, 1982; Simonyan, 2013), whereas indirect connectivity mediated through the brainstem reticular formation predominates in non-human primates (Simonyan and Jürgens, 2002). However, most previous studies did not specifically target and differentiate findings from dLMC or vLMC, the local sub-regional populations that appear to have very particular roles in different larynx functions, or compare connectivity to specific laryngeal muscles.

Humans are unique among primates in our ability to learn the flexible control necessary to support vocal communication. There is growing evidence that precise control of laryngeal function was one of several evolutionary developments that ultimately led to human-specific speech abilities (Belyk and Brown, 2017; Brown et al., 2008; Fitch, 2000; Fitch et al., 2016; Ghazanfar and Rendall, 2008; Hickok, 2016; Pisanski et al., 2016). The dynamic ways that pitch is used to communicate complex linguistic meaning, as in our contrastive emphasis task, may reflect a specialization not present in other species. Outside of primates, however, skillful vocal control can be learned in a limited group of species including songbirds, parrots, cetaceans, bats, and elephants (Janik and Slater, 1997; Jarvis, 2004; Fitch et al., 2010). The central control of vocal pitch is well described in songbirds (Sober et al., 2008) and recent work suggests potential convergent genetic features between human LMC and songbird sensori-motor nuclei (Pfenning et al., 2014). These songbird nuclei also feature both auditory and motor representations, and this is thought to have significant implications

for mimicking behavior and vocal learning (Prather et al., 2008). Indeed, more comparative research with both anatomy and finegrained neurophysiology holds great promise to fully understand the unique specializations of the human LMC that supports the unique capacity for speech.

STAR Methods

Contact for Reagent and Resource Sharing

Further information and requests for the data used in this study should be directed to and will be fulfilled by the Lead Contact, Dr. Edward Chang (edward.chang@ucsf.edu).

Experimental Model and Participant Details

All 12 participants were native English-speaking patients who underwent chronic implantation of a subdural electrocorticography (ECoG) array as part of their surgical treatment of epilepsy. Six of the participants had ECoG grids on the left hemisphere, and six had ECoG grids on the right hemisphere. Seven were female and five were male. All participants were adult (>18 years of age; range: 18–54 years old). Electrode coverage for all patients is shown in Figure S1. We did not perform any analysis of the influence of sex, gender identity, or both on the results of the study given the limited sample size. Participants gave their written informed consent. Each participant reported normal speaking and hearing ability. The experimental protocol was approved by the Human Research Protection Program and the UCSF Institutional Review Board, which reviews the safety and ethics of human research studies.

Method Details

Neural recordings

Each participant was unilaterally implanted with a 256-channel lattice array of electrodes, each with an exposed diameter of 1.17 mm and center-to-center spacing of 4 mm. Cortical local field potentials were amplified and quantized using a pre-amplifier (PZ5, Tucker-Davis Technologies), and preprocessed using a digital signal processor (RZ2, Tucker-Davis Technologies).

Preprocessing

The voltage trace of each electrode was visually inspected for artifact and excessive noise, and noisy electrodes were excluded from further analysis. For the remaining electrodes, we down-sampled the signal to 400 Hz and used a common average reference across electrode blocks and notch filters at 60, 120 and 180 Hz to remove line noise. For each electrode, we extracted the time-varying high gamma (HG) analytic amplitude using eight Gaussian band pass filters with centers between 70 and 150 Hz (73.0, 79.5, 87.8, 96.9, 107.0, 118.1, 130.4, and 144.0 Hz) with increasing σ (4.68, 4.92, 5.17, 5.43, 5.70, 5.99, 6.30, and 6.62 Hz). We then used a Hilbert transform (Moses et al., 2016). HG was calculated as the mean of these bands, and z-score was computed relative to the entire experimental block.

Acoustic analysis

We extracted voicing and the pitch contour of each sentence with an autocorrelation method in Praat (Boersma, 1993). Pitch minimum and maximum were determined individually for each participant, and a timestep of 0.0025 was used. All other parameters were the Praat default parameters. The pitch contour was then post-processed. We used an 80 ms median filter, then corrected erroneous octave jumps, interpolated through unvoiced regions in log(Hz), and filtered with an 80 ms Hanning window. Throughout the text, “pitch” refers to fundamental frequency.

Intensity was also extracted from each trial using Praat and normalized by recording session. The intensity signal was calculated as the square of the signal convolved with a Gaussian window of length $3.2/(\text{minimum pitch})$ which was determined on individually for each participant.

We found that some articulatory features tended to be correlated with pitch. For instance, nasals tended to have low pitch in all prosodic styles, so electrodes that were strongly tuned to velar movements would appear to be negatively correlated with pitch. To address this potential confound, an acoustic model was used. Although the pitch contours varied, the same syllable sequence was spoken each time, allowing us to examine specifically the control of pitch during natural speech and control for the articulatory movement of the production of the syllables.

We calculated Mel frequency cepstral coefficients (MFCCs) to temporally align the sentences for each participant (McFee et al., 2015). A power spectrogram of the microphone signal was calculated using a short-time Fourier transform with a window length of 2048 and a hop length of 512. The frequencies of the spectrogram were mapped to the Mel scale using triangular overlapping windows. The power was converted to decibels and finally a discrete cosine transform was used, which resulted in the final MFCCs.

Dynamic time warping was used to align the sentences based on the MFCCs (Slavador and Chan 2007). One of the sentences was arbitrarily chosen as the template sentence. For all other sentences, time window segments were duplicated and/or removed to find the temporal “warp” that minimized the Euclidean distance between the MFCCs of that sentence and the template sentence. This warp was then applied to the pitch, intensity, and high gamma analytic amplitude of each sentence. By removing the average neural activation across trials in this new timing, we removed the contribution of neural representation of articulatory movements that were consistent across trials.

This acoustic model does not track the articulators directly (Bouchard et al., 2016) or explicitly model the movement of specific articulators from the acoustics (Bouchard et al., 2013), but implicitly models the supra-laryngeal articulators by their effect on the acoustics of the sentence. Our approach has the advantage of being free of modeling assumptions about the relationship between neural activation and articulator movement (e.g. linearity). However, it does not capture trial-to-trial differences in articulation beyond timing differences. For instance, if a participant dropped the “r” of “never” for one trial, an explicit model might capture this but our approach would not. We expect these differences to be

relatively rare and small for our task, where the syllabic context is the same across repetitions.

Fujisaki Parameter Estimation

The Fujisaki model of vocal pitch is a model that separates the pitch contour of an utterance (F_0) into three components, the phrase (P), the accent (A_c), and the baseline (F_b). The phrase is composed of I individual phrase gestures of amplitude $A_{p,i}$ and shape G_p . The accent is composed of J individual accent gestures of amplitude $A_{a,j}$ and shape G_a . The model is defined by the following equations (Fujisaki, 2004):

$$\ln F_0(t) = \ln F_b + P + A_c$$

$$P = \sum_{i=1}^I A_{p,i} G_p(t - T_{0i}) \quad A_c = \sum_{j=1}^J A_{a,j} \{G_a(t - T_{1j}) - G_a(t - T_{2j})\}$$

$$G_p(t) = \begin{cases} \alpha^2 t e^{-\alpha t}, & t \geq 0, \\ 0, & t < 0 \end{cases} \quad G_a(t) = \begin{cases} \min[1 - (1 + \beta t)e^{-\beta t}, \gamma], & t \geq 0, \\ 0, & t < 0 \end{cases}$$

The phrase and accent components were estimated for each spoken sentence using FujiParaEditor (Mixdorff, 2000). We used automated inference (Mixdorff, 2009), with manual corrections where necessary.

Singing

The singing performance was measured quantitatively for each participant. First, the value of each note was determined by the median pitch produced for the duration of the note. To enable comparison between participants with different vocal ranges, each note converted to a semitone value:

$$s = 12 \log_2(f)$$

where f is pitch and s is semitone. Using the semitone values, the performance of each singer was measured by the average interval between “do” and “sol” (target = 7.0) and the standard deviation for low and high notes. A single participant was best in both of these metrics (black, Figure S3), and is used as the example participant in Figure 3. This participant also had approximately the same loudness distribution for low and high notes. A cross-participant analysis was used on the remaining participants. Several of these participants were not able to successfully mimic the melody of the task, but were still able to sing notes that varied in pitch.

Stimulation Mapping

Intraoperative direct electrical stimulation mapping of the peri-rolandic cortices was performed in 18 participants (5 left) as a part of their clinical care prior to surgical resection (4 of these participants also participated in the contrastive emphasis and singing task

experiments). After the induction of anesthesia, electromyography needles were placed in the orbicularis oris, tongue, and hand by a certified neuromonitoring specialist. A NIM® endotracheal tube (Medtronic, Minneapolis, MN) was placed under direct visualization with wire electrodes in contact with the vocal folds bilaterally to record laryngeal EMG activity (Eisele, 1996; Rea and Khan, 1998). The time-locked EMG activity and stimulation parameters were recorded on a Cascade® intraoperative neuromonitoring system (Cadwell, Kennewick, WA). The use of the NIM endotracheal tube was originally developed for monitoring laryngeal nerve function during neck surgeries. In our practice, it has been added to our routine motor mapping protocols because it is non-invasive, adds no additional risk, is very reliable compared to monitoring face movements, and permits the monitoring of vocal muscle EMG which is not visible or detectable otherwise. No adverse events have been encountered with its use over the past five years.

A craniotomy was performed, the dura was opened, and the exposed fronto-temporo-parietal cortical surface was densely mapped. The mapping was performed using a bipolar Ojemann Cortical Stimulator® probe (Integra, Plainsboro, NJ) with 5mm electrode spacing. The stimulator probe was applied sequentially to one cortical site at a time, as the voltage was increased from 0V to 100V, in increments of 5–10V, or until an EMG response was observed at that site. A train of 5–9 biphasic square waves, each with equal positive and negative phases of 75µs duration was used (Tate et al., 2013) For each trial of stimulation, the voltage was held constant, while the current was allowed to vary. EMG activity was simultaneously recorded from orbicularis oris, tongue, hand, and larynx as voltage was increased on each trial at each cortical site. Sites of cortical stimulation were spaced approximately 3–5mm apart. If an evoked potential was observed from any of the EMG electrodes, a voltage threshold was identified and the corresponding cortical site was photographed and recorded on the participant's coregistered MRI surface reconstruction using the BrainLab® neuronavigation system. The cortical sites from each participant were then warped into a common space for visualization (see previous description of electrode warping, Hamilton et al, 2017). Relative localization of the arm and mouth were determined by normalizing the location of the sites of each participant to the dorsal-most laryngeal site.

In 4 right hemisphere participants, multiple additional trials of stimulation across a range of voltages was performed at the dLMC site evoking laryngeal EMG activity, in order to characterize the relationship between dLMC stimulation voltage and the magnitude of laryngeal muscle activation. The cortical site was stimulated at voltages ranging from 10–15V below threshold, up to 100V or the plateau of the laryngeal EMG response. All stimulations were performed 5–10 seconds apart to avoid adaptation. EMG voltage responses were filtered with an 8th order Butterworth filter with critical frequency 32 Hz. The normalized peak-to-trough amplitude of the motor evoked potentials recorded from the laryngeal EMG was plotted as a function of stimulation voltage. Normalization was relative to the range of peak-to-trough response for each vocal fold of each participant.

In an independent cohort of patients undergoing craniotomy for surgical resection in the left, dominant hemisphere, stimulation mapping was performed with the patients fully conscious and conversant in order to identify speech areas (see previously published awake mapping protocol) (Breshears et al., 2015; Chang et al., 2016). After exposure of the peri-rolandic

cortex and emergence from intravenous sedation (either dexmedetomidine or propofol), intravenous fentanyl was titrated for optimal balance of pain control with patient arousal during the mapping procedure. The exposed cortex was densely mapped using an Ojemann stimulator (current range: 1 to 3.5 milliamps, pulse frequency 60Hz, pulse width 1ms, stimulus duration: 500 to 1500ms, stimulator electrode spacing: 5 mm). Prior to mapping, the after-discharge threshold was determined; the mapping was conducted at the maximum current that did not result in cortical spread (i.e. after-discharges). This ensured a low false negative rate. Each response or nonresponse to stimulation was tested for consistency/repeatability with at least 3 nonconsecutive stimulations. Responses were considered valid only in the absence of afterdischarges or seizure activity on electrocorticography, which was monitored and reported in real-time by an epileptologist. The mapping procedure was recorded simultaneously with 2 video cameras, one with an unobstructed view of the patient's face, and the second with an unobstructed view of the cortical surface. Cortical sites evoking involuntary vocalization responses were documented with a photograph and transferred onto the patient's cortical surface reconstruction from their MR imaging. These were warped into a common space, as described above. Acoustic waveforms of the vocalizations were extracted from the audio files for spectral analysis using librosa (McFee et al., 2015).

Quantification and Statistical Analysis

Pitch Tuning

After using dynamic time warping to remove the representation of the syllabic structure of the sentence (see Methods Details: Acoustic Analysis), we used a trial-wise permutation test to test the significance of the Pearson correlation between the neural activation and pitch (Figure 1C). The neural activation was shuffled with respect to the pitch contours (number of permutations = 1000, p-value < 0.001).

To determine the functional relationship between pitch and neural activation pitch was digitized into 20 bins uniformly spanning the middle 90-percentile range of pitch values for each participant, and the mean and standard deviation of high-gamma was calculated for each bin and significant electrode (Figure 1D).

To calculate the timing of the neural response in speech and in listening, we quantified the relative activation time of each significant electrode as the time when the high gamma analytic amplitude crosses a 1 s.d. threshold relative to the acoustic onset of the sentence. We used a paired t-test to determine a significant difference between the relative activation times in speech vs. listening (number of electrodes = 12, p-value < 0.001).

Fujisaki Parameter Selectivity

Correlation of neural activity (z-scored high gamma analytic amplitude) was calculated against P and A_c and against the binary voicing metric (V) extracted from Praat. For each metric and each of the 3,072 electrodes, we conducted a permutation test similar to the significance test for pitch (number of permutations = 1000, $p < 0.001$). Since the activation of many electrodes were correlated weakly with phrase due to sentence timing, electrodes

were also required to have Pearson correlation > 0.2 to be labeled significantly tuned with P (Figure 2G).

Singing

The median pitch through the duration of each note was used as the note's pitch, and we calculated a timepoint-by-timepoint correlation between high-gamma activation and pitch for each electrode in the dLMC. Figure 3D shows a statistically significant Pearson correlation (number of electrodes = 62, p -value < 0.01) between the encoding of pitch for the same electrodes during speaking and during singing. In this case, no dynamic time warping or partial correlation with intensity was conducted so that the correlations could be compared directly.

Stimulation Mapping

For the four anesthetized participants that received graded stimulation, a significant Pearson correlation was calculated between the voltages of the stimulations and the normalized laryngeal response magnitudes ($n = 184$ stimulations, Pearson $r = 0.85$, p -value $< 1e-52$). For the example participant, two one-tailed 1-sample t -tests were conducted testing difference between the higher and lower extreme values (0 and 100 V) and the repeated 60 V stimulus (t -test, $n=11$, $p<0.01$) (Figure 4C).

Data and Software Availability

All data and code are available upon request to the Lead Contact.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Ken Probst for anatomy illustrations in Figure 1. We thank Kristina Simonyan, Keith Johnson, and John Houde for their helpful comments on the manuscript. This work was supported by grants from the NIH (DP2 OD008627 and U01 NS098971-01). E.F.C is a New York Stem Cell Foundation-Robertson Investigator and Bowes Biomedical Investigator. This research was also supported by The New York Stem Cell Foundation, the Howard Hughes Medical Institute, The McKnight Foundation, The Shurl and Kay Curci Foundation, and The William K. Bowes Foundation.

References

- van Alphen P , and van Bergem DR (1989). Markov models and their application in speech recognition, Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam13: 1–26.
- Behroozmand R , Shebek R , Hansen DR , Oya H , Robin DA , Howard MA , and Greenlee JDW (2015). Sensory – motor networks involved in speech production and motor control: An fMRI study. *Neuroimage* 109, 418–428.25623499
- Belyk M , and Brown S (2015). Pitch underlies activation of the vocal system during affective vocalization. *Soc. Cogn. Affect. Neurosci*
- Belyk M , and Brown S (2017). The origins of the vocal brain in humans. *Neurosci. Biobehav. Rev* 77, 177–193.

- Boersma P (1993). Accurate Short-Term Analysis of the Fundamental Frequency and the Harmonics-To-Noise Ratio of a Sampled Sound. *Proc. Inst. Phonetic Sci* 17, 97– 110.
- Bouchard KE , Mesgarani N , Johnson K , and Chang EF (2013). Functional organization of human sensorimotor cortex for speech articulation. *Nature*.
- Bouchard KE , Conant DF , Anumanchipalli GK , Dichter B , Chaisanguanthum KS , Johnson K , and Chang EF (2016). High-resolution, non-invasive imaging of upper vocal tract articulators compatible with human brain recordings. *PLoS One* 11, 1–30.
- Breshears JD , Molinaro AM , and Chang EF (2015). A probabilistic map of the human ventral sensorimotor cortex using electrical stimulation. *J. Neurosurg* 123, 3–5.
- Brown S , Ngan E , and Liotti M (2008). A larynx area in the human motor cortex. *Cereb Cortex* 18, 837–845.17652461
- Chang EF , Niziolek C. a , Knight RT , Nagarajan SS , and Houde JF (2013). Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proc. Natl. Acad. Sci. U. S. A* 110, 2653–2658.23345447
- Chang EF , Breshears JD , Raygor KP , Lau D , Molinaro AM , and Berger MS (2016). Stereotactic probability and variability of speech arrest and anomia sites during stimulation mapping of the language dominant hemisphere. *J. Neurosurg* 126, 1–4.26894456
- Cheung C , Hamiton LS , Johnson K , and Chang EF (2016). The auditory representation of speech sounds in human motor cortex. *Elife* 5, 1–19.
- Collier R (1975). Physiological correlates of intonation patterns. *The Journal of the Acoustical Society of America*, 58(1), 249–255.1150974
- Conant DF , Bouchard KE , Leonard MK , & Chang EF (2018). Human sensorimotor cortex control of directly-measured vocal tract movements during vowel production. *Journal of Neuroscience*, 2382–17.
- Crone NE , Miglioretti DL , Gordon B , and Lesser RP (1998). Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Eventrelated synchronization in the gamma band. *Brain* 121(12), 2301–2315.9874481
- Eisele DW (1996). Intraoperative electrophysiologic monitoring of the recurrent laryngeal nerve. *Laryngoscope* 106, 443–449.8614219
- Fitch WT (2000). The evolution of speech: a comparative review. 6613, 258–267.
- Fitch WT , Huber L , Bugnyar T (2010) Social cognition and the evolution of language: constructing cognitive phylogenies. *Neuron* 65: 795–814.20346756
- Fitch WT , Boer B. De , Mathur N , and Ghazanfar AA (2016). Monkey vocal tracts are speech-ready.
- Foerster O (1936). *Motorische Felder und Bahnen*, Handbuch Neurologie, Bumke-Foerster Bd.
- Fujisaki H (2004). *Speech Prosody 2004 Information, Prosody, and Modeling*.
- Ghazanfar AA , and Rendall D (2008). Evolution of human vocal production. *Curr. Biol* 457–460.
- Guenther FH (2006). Cortical interactions underlying the production of speech sounds. *J. Commun. Disord* 39, 350–365.16887139
- Hast MH , and Milojkvic R (1966). The response of the vocal folds to electrical stimulation of the inferior frontal cortex of the squirrel monkey. *Acta Otolaryngol.* 61, 196–204.4959999
- Hickok G (2016). A cortical circuit for voluntary laryngeal control : Implications for the evolution language. *Psychon. Bull. Rev*
- Hull DM (2013). Thyroarytenoid and cricothyroid muscular activity in vocal register control.
- Janik VM , Slater PJB (1997) Vocal learning in mammals. *Adv Stud Behav* 26: 59– 99.
- Jarvis ED (2004) Learned birdsong and the neurobiology of human language. *Ann NY Acad Sci* 1016: 749–777.15313804
- Jürgens U (1974). On the elicibility of vocalization from the cortical larynx area. *Brain Res.* 81, 564–566.4215545
- Jürgens U (2002). Neural pathways underlying vocal control. *Neurosci Biobehav Rev* 26, 235–258.11856561
- Jürgens U (2009). The Neural Control of Vocalization in Mammals: A Review. *J. Voice* 23, 1–10.18207362

- Kirzinger A , & Jürgens U (1982). Cortical lesion effects and vocalization in the squirrel monkey. *Brain research*, 233 (2), 299–315.7059813
- Kuypers HG : Corticobular connexions to the pons and lowerbrain-stem in man: an anatomical study. *Brain* 1958,81:364–388.13596471
- Ladd DR (1984). Declination: a review and some hypotheses. *Phonology* 1, 53–74.
- Ladd DR (2008). *Intonational phonology*. Cambridge University Press.
- Lieberman PH , Klatt DH , & Wilson WH (1969). Vocal tract limitations on the vowel repertoires of rhesus monkey and other nonhuman primates. *Science*, 164(3884), 1185– 1187.4976883
- Loucks TM , Poletto CJ , Simonyan K , Reynolds CL , & Ludlow CL (2007). Human brain activation during phonation and exhalation: Common volitional control for two upper airway functions. *Neuroimage*, 36(1), 131–143.17428683
- Ludlow CL (2005). Central nervous system control of the laryngeal muscles in humans. *Respiratory Physiology & Neurobiology*, 147(2–3), 205–222. 10.1016/j.resp.2005.04.01515927543
- Mayer J , Wildgruber D , Riecker A , Dogil G , Ackermann H , and Grodd W (2002). Prosody production and perception: converging evidence from fMRI studies. *Proc. Int. Conf. Speech Prosody* 487–490.
- McFee B , Colin R , Dawen L , Daniel EPW , Matt M , Eric B , and Oriol Nieto. (2015) librosa: Audio and music signal analysis in python In *Proceedings of the 14th python in science conference*, pp. 18–25. DOI: 10.5281/zenodo.293021.
- Mixdorff H (2000). A Novel Approach to the Fully Automatic Extraction of Fujisaki Model Parameters. *Proc. Int. Conf. Acoust. Speech Signal Process.* 1, 1281–1284.
- Mixdorff H (2009). FujiParaEditor. <http://public.beuth-hochschule.de/~mixdorff/thesis/fujisaki.html>
- Moses DA , Mesgarani N , Leonard MK , and Chang EF (2016). Neural speech recognition: continuous phoneme decoding using spatiotemporal representations of human cortical activity. *J. Neural Eng.* 13, 56004.
- Nishida M , Korzeniewska A , Crone NE , Toyoda G , Nakai Y , Ofen N , Brown EC , and Asano E (2017). Brain network dynamics in the human articulatory loop. *Clin. Neurophysiol* 128, 1473–1487.28622530
- Niziolek CA , Nagarajan SS , and Houde JF (2013). What Does Motor Efference Copy Represent? Evidence from Speech Production. *J. Neurosci* 33, 16110–16116.24107944
- Olthoff A , Baudewig J , Kruse E , and Dechent P (2008). Cortical sensorimotor control in vocalization: A functional magnetic resonance imaging study. *Laryngoscope* 118, 2091–2096.18758379
- Patira R , Ciniglia L , Calvert T , & Altschuler EL (2017). Pure apraxia of speech due to infarct in premotor cortex. *Neurologia i neurochirurgia polska*, 51(6), 519–524.28844393
- Penfield W , and Boldrey E (1937). Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain* 60, 389–443.
- Penfield W , and Roberts L (1959). *Speech and brain mechanisms* (Princeton).
- Pfenning AR , Hara E , Whitney O , Rivas MV , Wang R , Roulhac L , Howard JT , Wirthlin M , Lovell PV , Mouncastle J , et al. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science* 346(6215), 1256846.25504733
- Pisanski K , Cartei V , McGettigan C , Raine J , and Reby D (2016). Voice Modulation: A Window into the Origins of Human Vocal Control? *Trends Cogn. Sci* xx, 1–15.
- Prather JF , Peters S , Nowicki S , and Mooney R (2008). Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature* 451, 305–310.18202651
- Press WH , Flannery BP , Teukolsky SA , and Vetterling WT (1989). *Numerical Recipes*, Cambridge University Press.
- Rathelot J , and Strick PL (2009). Subdivisions of primary motor cortex based on cortico-motoneuronal cells. *PNAS* 106.
- Ray S , and Maunsell JHR (2011). Different origins of gamma rhythm and highgamma activity in macaque visual cortex. *PLoS Biol.* 9.

- Rea JL , and Khan A (1998). Clinical evoked electromyography for recurrent laryngeal nerve preservation: use of an endotracheal tube electrode and a postcricoid surface electrode. *Laryngoscope* 108, 1418–1420.9738770
- Rodel RM , Olthoff A , Tergau F , Simonyan K , Kraemer D , Markus H , and Kruse E (2004). Human cortical motor representation of the larynx as assessed by transcranial magnetic stimulation (TMS). *Laryngoscope* 114, 918–922.15126757
- Rooth M (1992). A theory of focus interpretation. *Nat. Lang. Semant. An Int. J. Semant. Its Interfaces Gramm.* 1, 75–116.
- Roubeau B , Chevrie-Muller C , & Saint Guily JL (1997). Electromyographic activity of strap and cricothyroid muscles in pitch change. *Acta oto-laryngologica*, 117(3), 459–464.9199535
- Salvador S , and Chan P (2007) FastDTW: Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis* 11.5 561–580.
- Sammler D , Anwender A , Bestelmeyer PEG , and Belin P (2015). Dorsal and Ventral Pathways for Prosody. 3079–3085.
- Scherer KR (1989). Vocal correlates of emotional arousal and affective disturbance In *Handbook of Psychophysiology: Emotion and Social Behavior*, (London: John Wiley & Sons), pp. 165–197.
- Simonyan K (2013). The laryngeal motor cortex: its organization and connectivity. *Current opinion in neurobiology*, 28, 15–21.
- Simonyan K , and Horwitz B (2011). Laryngeal Motor Cortex and Control of Speech in Humans. *Neuroscientist* 17, 197–208.21362688
- Simonyan K , and Jürgens U (2002). Cortico-cortical projections of the motorcortical larynx area in the rhesus monkey. *Brain Res.* 949, 23–31.12213296
- Simonyan K and Jürgens U (2003). Efferent subcortical projections of the laryngeal motorcortex in the rhesus monkey. *Brain Res.* 974, 43–59.12742623
- Sober SJ , Wohlgenuth MJ , and Brainard MS (2008). Central contributions to acoustic variation in birdsong. *J. Neurosci* 28, 10370–10379.18842896
- Stevens SS (1935). The relation of pitch to intensity. *J. Acoust. Soc. Am* 6, 150–154.
- Tate MC , Guo L , McEvoy J , and Chang EF (2013). Safety and efficacy of motor mapping utilizing short pulse train direct cortical stimulation. *Stereotact. Funct. Neurosurg* 91, 379–385.24108152
- Tang C , Hamilton LS , & Chang EF (2017). Intonational speech prosody encoding in the human auditory cortex. *Science*, 357(6353), 797–801.28839071
- Titze IR , Luschei ES , and Hirano M (1989). Role of the thyroarytenoid muscle in regulation of fundamental frequency. *J. Voice* 3, 213–224.
- Wilson SM , Lam D , Babiak MC , Perry DW , Shih T , Hess CP , ... & Chang EF (2015). Transient aphasias after left hemisphere resective surgery. *Journal of neurosurgery*, 123(3), 581–593.26115463
- Wilson SM , Saygin AP , Sereno MI , and Iacoboni M (2004). Listening to speech activates motor areas involved in speech production. *Nat. Neurosci* 7, 701–702.15184903
- Wrench A (1999). MOCHA-TIMIT. Department of Speech and Language Sciences, Queen Margaret University College, Edinburgh, speech database.
- Zatorre RJ , and Baum SR (2012). Musical melody and speech intonation: Singing a different tune. *PLoS Biol.* 10, 5.

Highlights

The control of vocal pitch in the human laryngeal motor cortex CELL-D-18-00068R2

- A human brain area that controls vocal pitch in both speech and song is identified
- Two laryngeal functions, voicing and pitch, are encoded by distinct neural populations
- A causal role for larynx muscle control is demonstrated through cortical stimulation

In Brief:

The ability to control vocal pitch during speech and singing is encoded by the dorsal laryngeal motor cortex in humans

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

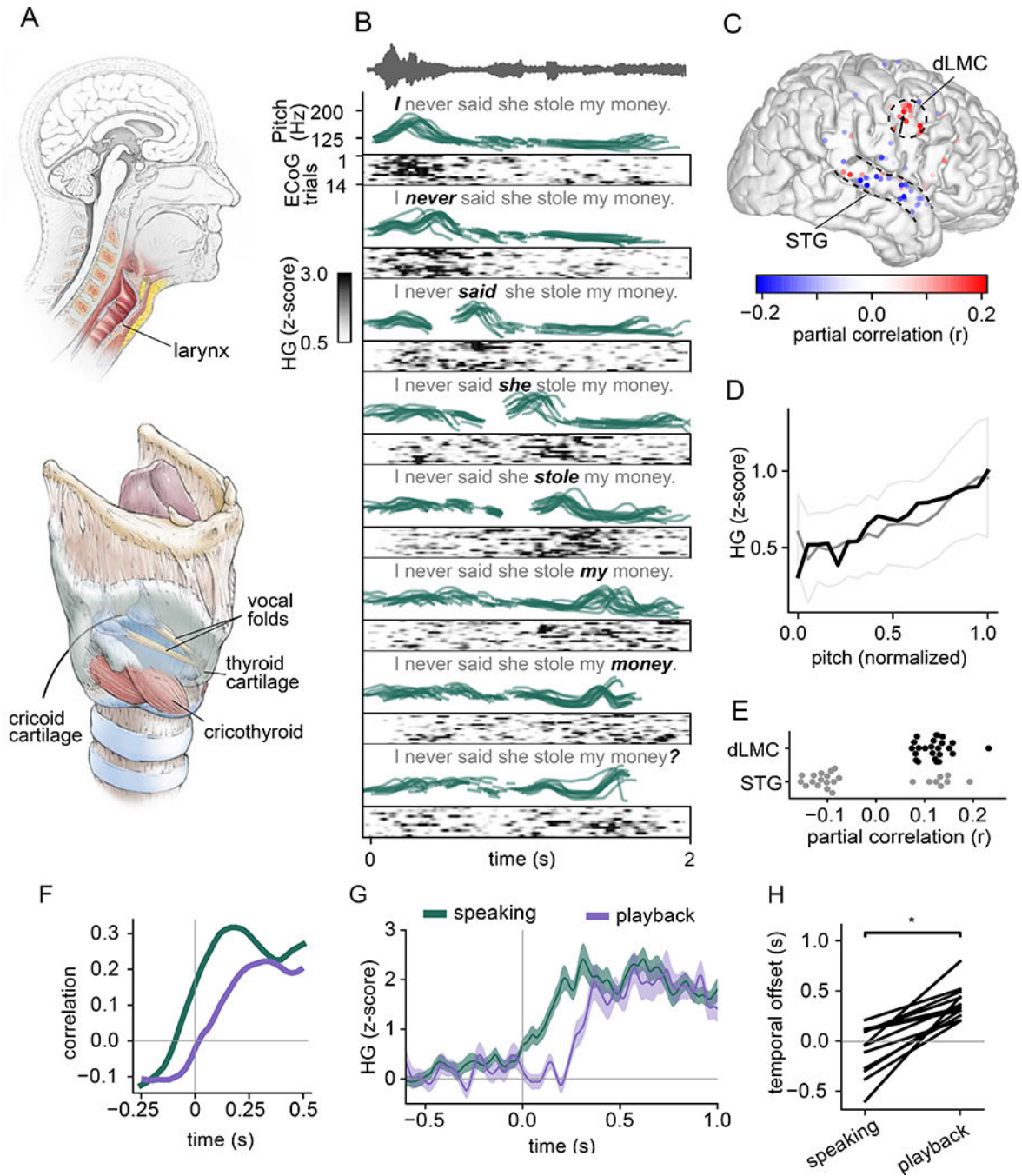


Figure 1 | Human cortical encoding of produced pitch in dLMC during a word emphasis task. Participants were instructed to emphasize specific words in a sentence. (A) Laryngeal anatomy. The vocal folds are stretched by the cricothyroid muscle, and increased tension in the vocal folds results in a higher produced pitch. (B) Pitch-correlated neural activity at an example electrode. The speech waveform for one example sentence (emphasis on "I") is shown at the top. Pitch contours (green lines) and single trial high gamma activation for the example electrode (black rasters) for every sentence spoken by a single participant are shown. Trials are grouped by the word of

emphasis and co-aligned to the beginning of the emphasized word. On a single trial level, transient increases in neural activity are associated with pitch change.

(C) Spatial localization of electrodes that have a significant correlation with pitch, after controlling for supralaryngeal articulators. Electrodes cluster on the anterior aspect of the precentral gyrus in the dorsal laryngeal motor cortex (dLMC, located lateral to hand and medial to the lip cortical representations). The right hemisphere is shown, and the arrow indicates the example electrode in (B). We also observed feedback responses in parabelt auditory cortex on superior temporal gyrus (STG).

(D) Relationship between pitch and high gamma (HG) cortical activity across all significant electrodes in dLMC (mean and s.d. in grey, example electrode in black) over normalized pitch range. Activation increases monotonically with pitch values (middle 90 percentile range plotted).

(E) Correlation values for significant electrodes in the dLMC and auditory STG regions. Electrodes in dLMC were all positively correlated with the produced pitch of the emphasized word, whereas activity of STG electrodes were both positively and negatively correlated with pitch.

(F-H) dLMC activity shows both motor and auditory response properties. Temporal analyses show dLMC activity during speaking conditions precedes playback (listening) of same utterances.

(F) Pearson cross-correlation for the example electrode in (B) for speaking (green) and playback (purple).

(G) Neural activation aligned to sentence onset for speaking (green) and playback (purple) conditions for the example electrode (mean \pm SEM).

(H) Average temporal offset of neural activation for each electrode in the dLMC with respect to sentence onset.

See also Figure S1.

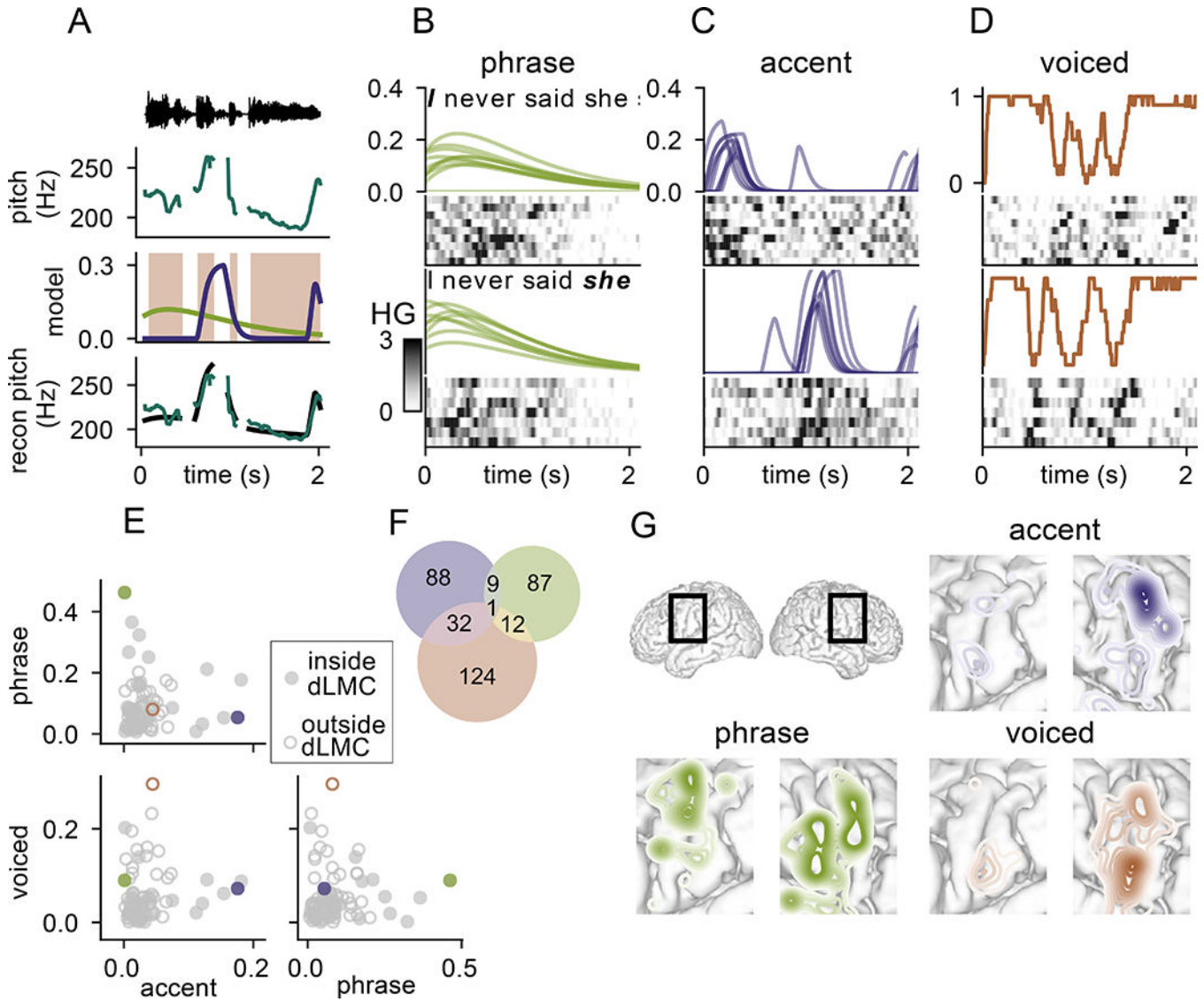


Figure 2 | Cortical representation of pitch contour components in speech: accent, phrase, and voicing.

(A) The Fujisaki model decomposes the pitch contour in natural speech into accent, phrase, and voicing components. Inference of the Fujisaki model is shown on an example sentence. In order from top to bottom: acoustic waveform of produced sentence; pitch contour extracted from sentence; phrase (green), accent (purple), and voicing (brown) components extracted from the pitch contour; original pitch contour (green) and Fujisaki reconstruction of pitch contour (black).

(B) Single-trial high gamma raster for an electrode controlling the phrase component of the pitch contour. Green curves show the phrase component of the Fujisaki model for each trial, and the grey rasters show the activation of an example “phrase” electrode ($r=0.45$). This electrode responded similarly to sentences with different accents (top and bottom).

(C) Single-trial high gamma raster for an electrode controlling pitch accents. Purple lines show the accent component for an example participant separated by sentence style, and the grey raster shows the activation of an example “accent” electrode ($r=0.17$).

(D) Single-trial high-gamma raster for an electrode controlling voicing. Brown lines show the proportion of sentences that are voiced for each style at each timepoint. This electrode has higher activation when the participant is voicing ($r=0.2$).

(E) The correlation coefficient between activation of the accent, phrase, and voicing components of the pitch model for each of the electrodes over the sensorimotor cortex.

Filled dots from inside and open dots from outside dLMC. Example electrodes in b-d are marked in their respective colors. Electrodes tend to be predominantly along the axes.

(F) Venn diagram showing numbers of electrodes with dissociable and joint encoding.

(G) Bilateral spatial location of electrodes on the vSMC across all participants. Accent and voicing electrodes were selected using a trial-wise shuffle test ($p<0.001$). Phrase electrodes were selected using a trial-wise shuffle test and a cutoff of $r<0.2$. Each brain reconstruction shows the kernel density estimation illustrating the spatial organization of electrodes on a common brain. Accent electrodes were strongly localized to the dLMC, while voicing and phrase electrodes were found in both the dLMC and the vLMC.

See also Figure S2.

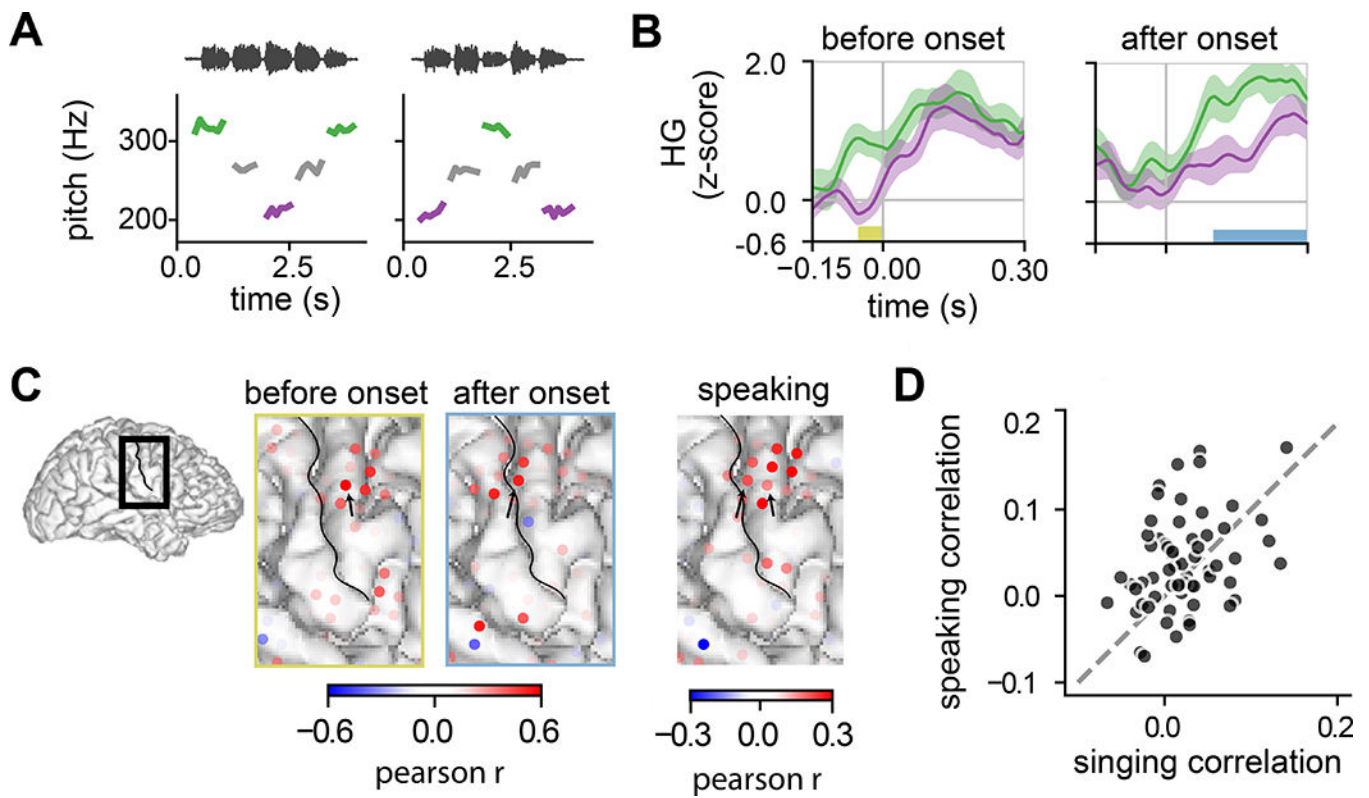


Figure 3 | Pitch encoding during singing.

(A) Singing task with two simple melodies. Notes are colored by low, middle, and high target tone. The sound waveforms are shown above, with produced pitch for each note below.

(B) High gamma response for two example electrodes in dLMC of the example participant for high (green) and low (purple) notes. Time=0 is the acoustic onset of the note. The yellow and blue segments mark time windows used to compute correlations in (C). Error bars are sem across trials.

(C) The Pearson correlation between cortical activation and vocal pitch for low and high notes using 50 ms before acoustic onset (left) and 100 – 300 ms after acoustic onset (middle). right: The Pearson correlation computed between pitch and high gamma activation for the contrastive emphasis task for this participant. Arrows mark the electrodes from (B), and the solid black line marks the central sulcus.

(D) Comparison between pitch encoding in dLMC electrodes during singing and during speaking for all participants. Pitch encoding was strongly correlated across electrodes in the two behavioral conditions (Pearson $r=0.33$, p -value < 0.01).

See also Figure S3.

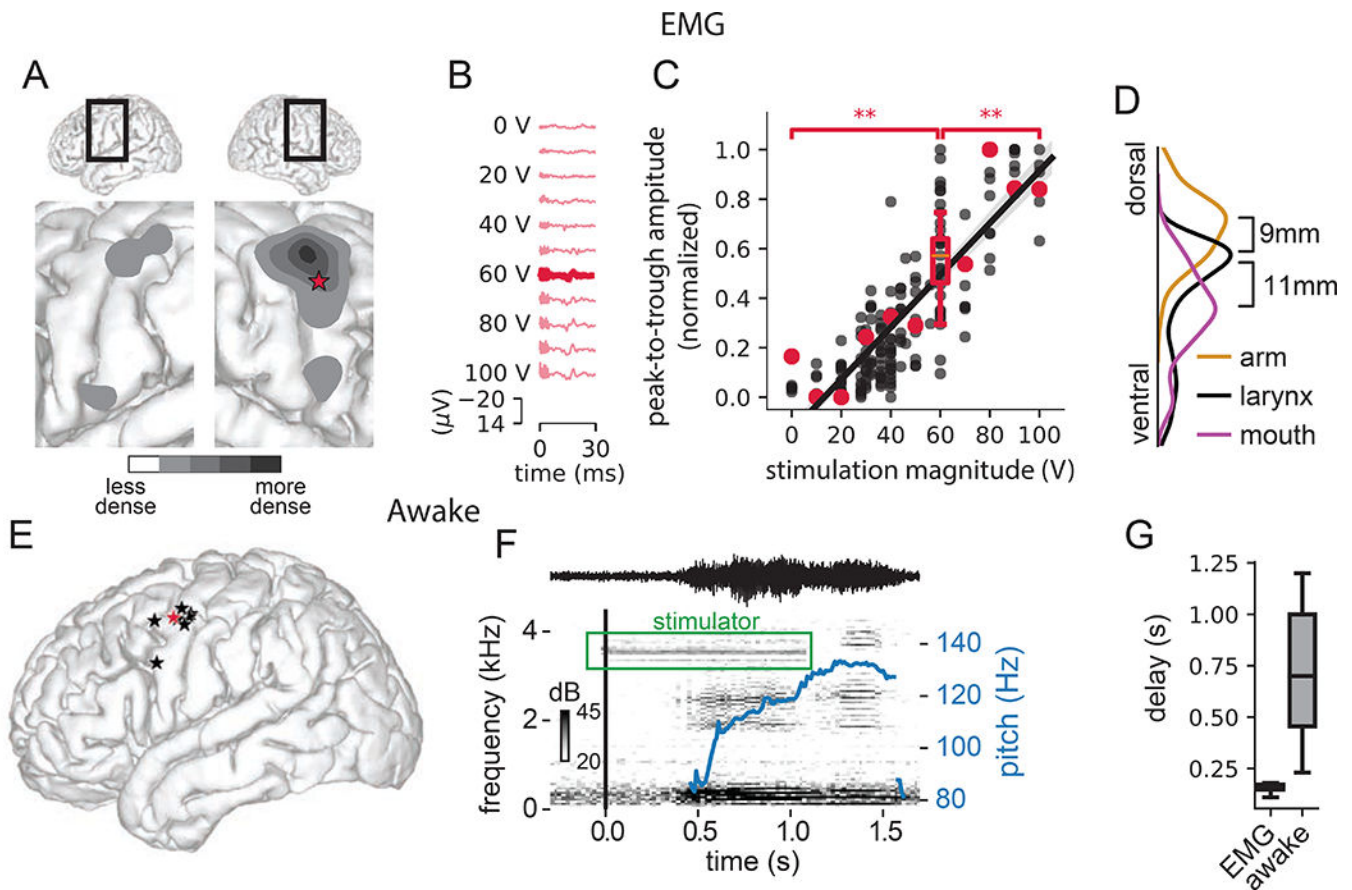


Figure 4 |. Electrical stimulation of dLMC.

(A) Cortical stimulation mapping of larynx responses in the primary sensory and motor cortices for 18 participants. The larynx was monitored using electromyography (EMG) electrodes on a customized endotracheal tube. The shading of the gray indicates the relative density of positive laryngeal response sites. Other evoked movements are not shown. The red star marks the example site that is shown in more detail in (B) and (C).

(B) Laryngeal EMG response for stimulations ranging from 0–100 V. Stimulation was delivered 11 times at 60 V and once at each of the other magnitudes for this patient.

(C) Three other patients also received graded stimulation. Peak-to-trough response amplitude was determined for each stimulation, and is shown for each patient, normalized to the maximum and minimum response for each larynx side of each participant. Responses to 60V are shown as a box plot (the borders are the range and the box edges are the quartiles). Laryngeal responses for the example stimulation site of (A) and (B) are shown in red. Stimulation responses to 60V are greater than 0V (p -value $< 1e-6$, one-sided t-test) and less than 100V (p -value $< 1e-3$, one-sided t-test). Therefore responses are not an all or none, but rather a graded response where more stimulation yields a greater laryngeal response. Stimulation magnitude is strongly correlated with laryngeal response across participants. The gray shading shows the standard error of the slope determined using bootstrapping ($n=1000$).

(D) Across participants, sites that evoked arm movement were dorsal of the larynx sites and sites that evoked mouth movement were ventral of the larynx sites.

(E) Sites that evoked a spontaneous involuntary voiced vocalization during awake stimulation mapping. The vocalization evoked by the red location is shown in (F).
(F) Spectrogram and pitch contour of an example evoked vocalization. Noise from the stimulator created a 3.5 kHz band in the spectrogram.
(G) Delay times between the start of stimulation and the beginning of the response for anesthetized (black) and awake (grey) stimulation. All of the response times for laryngeal response were shorter than times for vocalization response (the borders of the box plots mark the ranges and the box edges mark the quartiles).
See also Figure S4.