Auditory representation of learned sound sequences in motor regions of the macaque brain

Denis Archakov[a,b,1], Iain DeWitt[a,1], Paweł Kuśmierek[a,1], Michael Ortiz-Rios[a,2], Daniel Cameron[a,3], Ding Cui[a,4], Elyse L. Morin[a,5], John W. VanMeter[c], Mikko Sams[b], Iiro P. Jääskeläinen[b], and Josef P. Rauschecker[a,6]

[a]Department of Neuroscience, Georgetown University Medical Center, Washington, DC 20057; [b]Brain and Mind Laboratory, Department of Neuroscience and Biomedical Engineering, Aalto University School of Science, FI-02150 Espoo, Finland; and [c]Center for Functional and Molecular Imaging, Georgetown University Medical Center, Washington, DC 20057

Edited by Peter L. Strick, University of Pittsburgh, Pittsburgh, PA, and approved May 7, 2020 (received for review September 9, 2019)

Human speech production requires the ability to couple motor actions with their auditory consequences. Nonhuman primates might not have speech because they lack this ability. To address this question, we trained macaques to perform an auditory–motor task producing sound sequences via hand presses on a newly designed device ("monkey piano"). Catch trials were interspersed to ascertain the monkeys were listening to the sounds they produced. Functional MRI was then used to map brain activity while the animals listened attentively to the sound sequences they had learned to produce and to two control sequences, which were either completely unfamiliar or familiar through passive exposure only. All sounds activated auditory midbrain and cortex, but listening to the sequences that were learned by self-production additionally activated the putamen and the hand and arm regions of motor cortex. These results indicate that, in principle, monkeys are capable of forming internal models linking sound perception and production in motor regions of the brain, so this ability is not special to speech in humans. However, the coupling of sounds and actions in nonhuman primates (and the availability of an internal model supporting it) seems not to extend to the upper vocal tract, that is, the supralaryngeal articulators, which are key for the production of speech sounds in humans. The origin of speech may have required the evolution of a "command apparatus" similar to the control of the hand, which was crucial for the evolution of tool use.

auditory cortex | motor cortex | putamen | internal models

Speaking is a highly complex motor skill, engaging the coordinated use of ∼100 muscles (1). The production of the intended speech sounds must be precise and reproducible to assure reliable decoding during speech perception. It should not come as a surprise, therefore, that speech production lags significantly behind perception during language acquisition and takes several years to reach perfection (2). Most amazingly, during this process, children learn how to speak by simply listening to speech, not by receiving instruction on how to move their articulators (3). Auditory feedback thus plays a pivotal role in learning to speak, as demonstrated by the difficulty deaf children experience in acquiring this ability (4). How auditory signals reach the motor system and where the interaction between auditory and motor signals takes place is still poorly understood.

Early neuroethological work (5) proposed that sensorimotor coupling between perception and action generally involves internal "reafference" (or feedback) of the sensory signal for interaction with motor signals. The idea was later taken up by modern motor control theory (6, 7) and contributed to the concept of "internal models," which compute sensorimotor transformations as part of feedback control systems (8, 9). Specifically, forward internal models (or "forward models") model the causal relationship between actions and their consequences, whereas inverse internal models (or "inverse models") implement the opposite transformations, from desired consequences to actions (10). The use of internal models dramatically increases the speed and accuracy of movement by decreasing reliance on slow sensory feedback from the periphery and by minimizing the resulting error signal.

For all of the above reasons, feedback control systems incorporating internal models are increasingly applied to the study of speech production and its development (see refs. 11 and 12 for review). Building on early, pioneering work (13–15), studies incorporating optimality principles (16), such as optimal state estimation (17, 18) and state feedback control (19–22), have been successful in capturing important dynamics of human vocal communication. In fact, the emergence of internal models in brain systems for auditory–motor processing could be considered one of the key events during primate evolution that ultimately enabled speech in humans (23, 24). Similar arguments can be

Significance

Nonhuman primates have a rich repertoire of species-specific calls, but they do not show vocal learning. The usefulness of nonhuman primate models for studies of speech and language has, therefore, been questioned. This study shows that macaques can learn to produce novel sound sequences with their hands by pressing levers on a keyboard. Using awake functional MRI, we find activation of motor cortex and putamen, when the monkeys are listening to the same sound sequences they had learned to produce. The results indicate that auditory–motor training in monkeys can lead to the coupling of auditory and motor structures of the brain. This paradigm may be able to serve as a model system for the evolution of speech in primates.

Author contributions: D.A., P.K., M.O.-R., M.S., I.P.J., and J.P.R. designed research; D.A., P.K., M.O.-R., D. Cui, E.L.M., and J.P.R. performed research; D.A., I.D., P.K., M.O.-R., D. Cameron, J.W.V., and J.P.R. analyzed data; and D.A., I.D., P.K., M.O.-R., D. Cameron, M.S., I.P.J., and J.P.R. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Data deposition: MRI data associated with the figures in this work have been deposited on PRIME-DE (http://fcon_1000.projects.nitrc.org/indi/PRIME/georgetown.html).

[1]D.A., I.D., and P.K. contributed equally to this work.

[2]Present address: Bioscience Institute, Newcastle University Medical School, Newcastle upon Tyne NE2 4HH, United Kingdom.

[3]Present address: Department of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, ON L8S 4K1, Canada.

[4]Present address: Laboratory for Neuro- and Psychophysiology, Department of Neuroscience, Katholieke Universiteit Leuven, 3000 Leuven, Belgium.

[5]Present addresses: Yerkes National Primate Research Center, Atlanta, GA 30329; and Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, GA 30329.

[6]To whom correspondence may be addressed. Email: rauschej@georgetown.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1915610117/-/DCSupplemental.

First published June 15, 2020.

made for music (25, 26), although children arguably do not acquire musical abilities with the same ease as they acquire speech.

Despite remarkable similarities with humans in the anatomical organization of their auditory cortical pathways (17, 27), nonhuman primates do not have speech (24). While monkeys have an auditory vocalization system for species-specific communication (28, 29), they display little or no vocal learning (30), and training them to produce their vocalizations volitionally is notoriously difficult (31, 32). These shortcomings in vocal production of nonhuman primates compared to humans have been attributed to differences in the neural control of the vocal apparatus: the existence of direct descending projections from primary motor cortex to brainstem nuclei controlling the larynx in humans but not monkeys (33, 34), or increases in the relative volumes of frontal and parietal association cortex correlated with increases in the volume of the hypoglossal nucleus, which controls the tongue in hominoids and humans (35).

A more fundamental limitation would be the absence, in nonhuman primates, of an internal model structure coupling the auditory system with the vocal–articulatory apparatus as a whole, including its supralaryngeal parts, like tongue, lips, and jaw (24). While the existence of a forward model in conjunction with an efference copy ("corollary discharge") has been shown in the auditory cortex of monkeys (36, 37), evidence is lacking for the presence of an inverse model in the auditory–motor system of nonhuman primates. Importantly, the absence of an inverse model would prevent the mapping of comparisons between intended and produced sounds from auditory to motor code, essential for the learning and updating of motor programs.

Where in the brain such an inverse model for auditory–motor coupling could be implemented is an open question, given a probable hierarchy of internal model structures for speech and language (23). While a location in early sensory and motor areas (posterior/caudal superior temporal gyrus [STG] and precentral gyrus, respectively) would seem most natural, other locations are conceivable. In the monkey, lateral prefrontal cortex (LPFC) could provide an ideal place, because this is where high-level auditory structures of the auditory ventral pathway interface with motor planning areas of the dorsal pathway (17, 27) and cross-stream interaction takes place (23). Following the definitions of Romanski et al. (27) and Rauschecker and Scott (17), which built on earlier definitions from the visual system (38, 39), we term the recipient zone of the ventral stream in monkeys as ventrolateral prefrontal cortex (VLPFC), and that of the dorsal stream as dorsolateral prefrontal cortex (DLPFC). Correspondingly, Brodmann area 44 (BA44) is part of the dorsal stream, whereas BA45 is part of the ventral stream. Neurons in VLPFC (BA45) respond to species-specific vocalizations (40); articulator movements can be evoked by microstimulation in BA44 (41, 42). Although its anatomical structure is complex (43, 44), the inferior frontal gyrus (IFG) in humans, and "Broca's region" in it (consisting of both BA44 and BA45), may be considered the human functional homolog of LPFC (45). It is in this region where the transformation between auditory–semantic and articulatory–syntactic code (BA45 and BA44, respectively) in human speech occurs (17, 46, 47) and a major inverse model facilitating vocal control might, therefore, be situated.

In contrast to their limitations in vocal control, macaques possess highly developed abilities for fine motor control of their hands and arms, which are instrumental for tool use (48, 49). These skills (and the ability to acquire them by learning) suggest the existence of internal models coupling the somatosensory and motor systems for nimble use of the forelimbs. To test whether nonhuman primates might possess internal models for sensory–motor control of their upper extremities more generally, including auditory–motor control (which seems to be largely absent for the control of their vocal tract), we designed a paradigm to study learned auditory–motor behavior in rhesus monkeys: Instead of trying to train the animals to control their vocal articulators, they were trained to use their upper extremities for sound production. We hypothesized that rhesus monkeys can learn to produce sound sequences by moving their arms and pressing levers with their hands and fingers under auditory control. Our expectation was that such auditory–motor learning will lead to formation of internal models involving auditory and motor regions of the cerebral cortex that could then be detected using functional neuroimaging techniques.

Not only does speech require the activation of specific articulators, it also does so in well-defined action sequences. Given the known involvement of the basal ganglia (BG) in sensorimotor sequence learning (50–54), we hypothesized that successful learning of auditory–motor sequences may also be associated with BG activation.

Results

Three rhesus macaques (Do, Ra, and Ch) were taught to produce a repeating tone sequence using a special-built keyboard device ("monkey piano"; Fig. 1*A*) consisting of four levers that, when pressed, each produced a different tone of the same timbre. Each monkey was trained to play a different individual sequence ("melody") consisting of eight tones and gathered extensive experience with performing the task (see *Materials and Methods* and Fig. 1 for a detailed description of the device, the task, and behavioral performance).

As any motor production of sound sequences involves the pairing of motor actions with their acoustic consequences, we expected the formation of neuronal assemblies reflecting the association of auditory and motor information in the brain of trained monkeys. To identify these auditory–motor regions, we performed whole-brain functional magnetic resonance imaging (fMRI) in two of the animals (Do and Ra) that reached all necessary criteria (see below). We measured brain activation evoked by auditory stimulation with the same sound sequences that they had learned to produce (self-produced [SP]). Thus, the animals were only listening and were not producing the SP sequences inside the scanner, so any activation found would be driven by the auditory input, not by motor performance. The absence of residual or subliminal movement while listening to the learned sequences was confirmed by electromyography (EMG) outside the scanner (*SI Appendix*, Fig. S1). In addition, we used two sets of control stimuli in the fMRI study: 1) sound sequences with an equal number of tones that the monkeys had been passively exposed to (by listening to the sound sequences as produced by one of the other monkeys) for a comparable number of times but had never produced (non–self-produced [NSP]), and 2) sound sequences that the monkeys had neither produced nor were ever exposed to (unfamiliar [UF]) (see *Materials and Methods* for further details). The predicted outcome of the fMRI study was that the neuronal assemblies linking SP sounds with actions producing these sounds would be specifically activated by the SP sequence, and would thus lead to SP-evoked fMRI signal being significantly higher than NSP- or UF-evoked activation.

During the fMRI scans, the monkeys were awake and restrained in a horizontal, MRI-compatible primate chair. Their attention level was controlled by having them maintain visual fixation while they waited for a target sound (white-noise [WN] burst), which then required a saccade to another location to receive a reward. As they were fixating and undergoing fMRI scanning, the animals were presented with the SP sequences or the control sequences (NSP or UF), or with silent trials in an interleaved fashion (see *Materials and Methods* for details). Monkeys Do and Ra mastered both the saccade task and the piano task, producing sequences with increasingly consistent timing over the course of training (Fig. 1 *B* and *C*). When catch trials were interspersed during training (with lever presses producing an unexpected sound or no sound at all), both animals

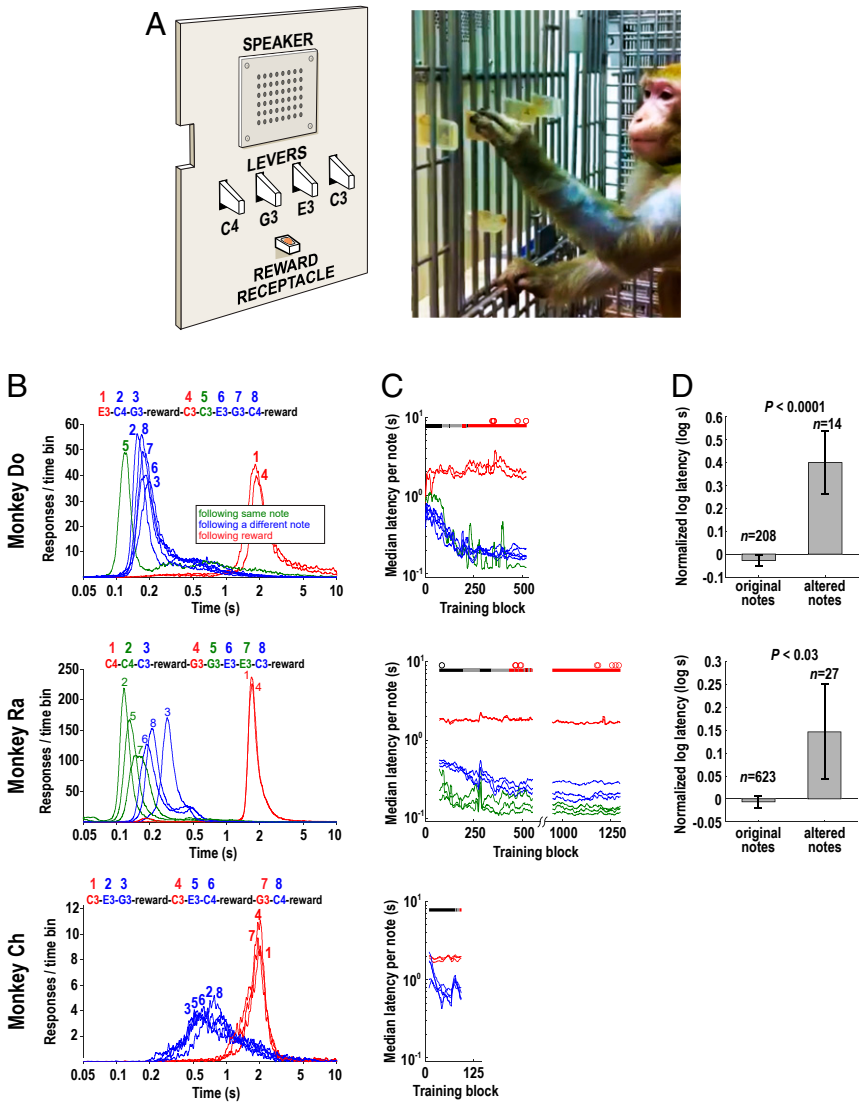

Fig. 1. The monkey piano apparatus and behavioral results. (*A*) Panel with four levers used for in-cage training (*Left*). A musical sound (C3, E3, G3, C4) was played upon corresponding lever press. (*Right*) Monkey Do performing the task. (*B*) Latency distributions of correct lever presses (relative to previous lever press) for each note in the sequence. Self-produced (SP) sequences shown at the *Top*. Depending on behavioral context, latencies fell into three types (color-coded; see *Inset*). Distribution numbers indicate the note's position in the sequence. (*C*) Progress of sequence training: median response latencies (see *B* for color-coding) in 50-sequence blocks. Black circles: scans performed early in sequence training (Ra only, *SI Appendix*, Fig. S7); red circles: scans in trained monkeys. The band below scan marks shows the amount of visual guidance: black, full (all eight presses in the sequence guided); gray, partial; red, final minimal. Latency data smoothed with a 10-block moving average. (*D*) Catch trials. Lever presses were programmed to occasionally produce a different sound. Latencies of lever presses following an altered note were significantly longer than following the original note. Error bars represent ±SEM. See *Materials and Methods* for details, and *SI Appendix*, Fig. S2.

altered their motor behavior, indicating that the monkeys were listening to the sounds during the performance and the sounds influenced their behavior (Fig. 1*D* and *SI Appendix*, Fig. S2).

We measured blood oxygenation level-dependent (BOLD) responses in the brain of both monkeys. All three sets of sound sequences (SP, NSP, and UF) robustly activated the inferior colliculus (IC) in the auditory midbrain as well as core and belt regions of auditory cortex (55, 56) on the superior temporal plane bilaterally in both monkeys ($q < 0.001$, false-discovery rate [FDR] corrected, cluster size $k \geq 10$ voxels, all stimulus conditions vs. baseline [silent trials]; Fig. 2). At lower thresholds, activation extended into parabelt regions of auditory cortex.

Most importantly, SP sound sequences selectively activated regions outside of classical auditory areas, specifically, motor regions of the cerebral cortex. The contrast SP vs. NSP and UF revealed significant activation centered in the left precentral gyrus in both animals (Do: $P < 0.01$, uncorrected; $k \geq 10$ voxels; Fig. 3 *A* and *B*; Ra: $q < 0.05$, FDR corrected; $k \geq 25$ voxels; Fig. 3 *E* and *F*). Surface rendering demonstrates that the precentral focus constitutes the global peak of cortical activation across the hemisphere for this comparison (Fig. 3 *A* and *E*). The same left precentral focus was found with separate contrasts of SP vs. NSP and SP vs. UF (*SI Appendix*, Fig. S3; see also Fig. 3 *C* and *G*). Amount of activation also correlated with behavioral performance (variability of lever press latencies) outside the scanner in monkey Ra (*SI Appendix*, Fig. S4). The activated region overlaps with primary motor cortex (referred to as area F1 or M1, depending on nomenclature) and extends into dorsal and ventral premotor areas F2 and F4 (57) (Fig. 3 *D* and *H*). Specifically, according to the foundational work of Matelli et al. (58,

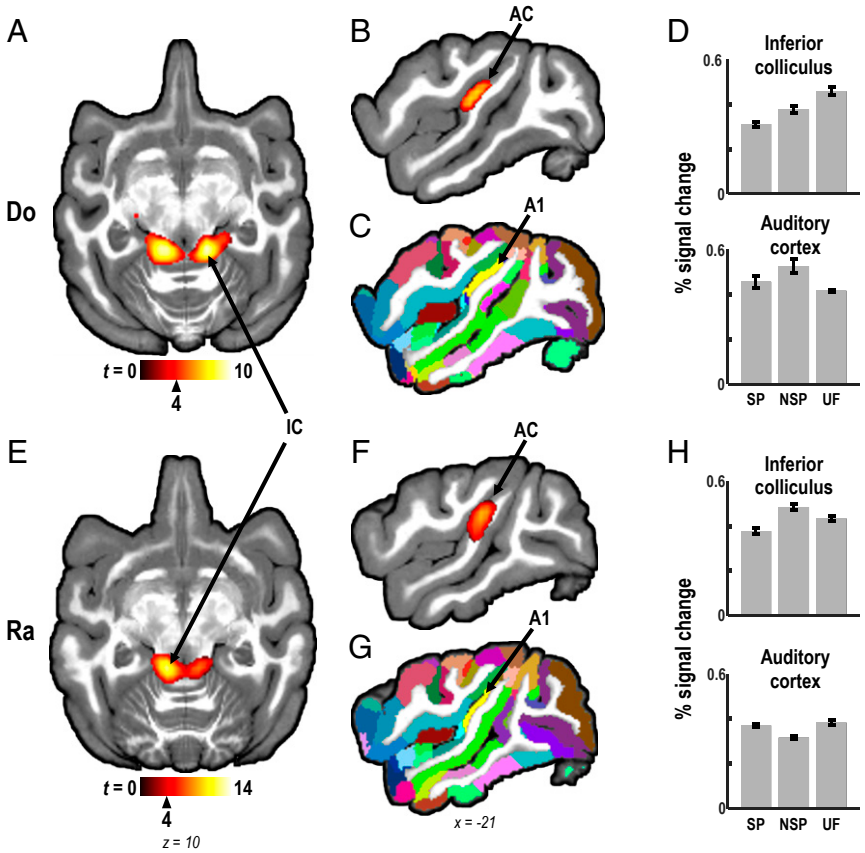

Fig. 2. Mapping of activation by sound sequences in the auditory pathway. Auditory stimuli (all experimental conditions combined, and compared to silence) evoked bilateral activation in auditory brainstem (inferior colliculus [IC]) (*A* and *E*) and in auditory cortex (AC) (core and belt) (*B* and *F*), including primary auditory cortex (A1), in both monkeys (Do and Ra; $q < 0.001$, FDR corrected; cluster size $k \geq 10$ voxels). (*C* and *G*) Color-coded D99 atlas segmentation (57), projected onto individual anatomy in the same monkeys, allowed assignment of BOLD activation to cortical areas. (*D* and *H*) AC and IC were reliably activated by stimuli of all three experimental conditions: self-produced sequences (SP), non–self-produced sequences (NSP), and unfamiliar sequences (UF). Bars show percent signal change relative to baseline within IC and AC foci defined by the "all stimuli vs. silence" contrast (mean ± SEM). In all figures showing MRI slices: Left side is shown on the left. Critical *t* values, associated with respective *q* and *P* values, for activation thresholds are shown next to the color bars with triangle markers.

59), the activated region (circled in Fig. 3*I*) represents the upper extremities.

Differential activation was also found for both animals in the BG, specifically the left putamen, for the contrast of SP vs. NSP and UF (Do: $P < 0.01$, uncorrected; $k \geq 10$ voxels; Ra: $q < 0.05$, FDR corrected; $k \geq 25$ voxels; Fig. 4), as well as for the SP vs. NSP and SP vs. UF contrasts analyzed separately (Fig. 4 *B* and *D* and *SI Appendix*, Fig. S5).

Discussion

Our behavioral data show that rhesus monkeys can successfully learn an auditory–motor task, producing sounds by pressing levers with their hands in a given sequence. The results of fMRI in the same monkeys demonstrate that auditory signals representing the sounds the monkeys learned to produce reach motor regions of the brain, confirming the existence of an inverse internal model. This experimental paradigm holds promise for the study of learned auditory–motor behaviors in primates and the evolution of speech and its neural basis.

The main question our study tried to address was why nonhuman primates, despite remarkable similarities with humans in the anatomical organization of their auditory cortical pathways (17, 60), including a well-established dual-pathways architecture in both species (27, 56, 61, 62), do not have a more speech-like (or song-like) communication system, including vocal learning and volitional control. The similarity between humans and monkeys is especially striking in terms of the hierarchical organization of the auditory ventral stream for the decoding of complex sounds, including species-specific vocalizations and speech (63–65). The dorsal stream, which contains direct projections connecting posterior STG and lateral/inferior PFC (as well as indirect connections via posterior parietal cortex) in both species (27), harbors spatial as well as sensorimotor functions (17). On the basis of diffusion tensor imaging, it was found that the direct projection is denser and more left-lateralized in humans than in monkeys (66, 67), but whether this quantitative difference can explain the qualitative difference in vocal production appears doubtful (68). Species differences in auditory–motor functions, including vocal behavior, are more likely to be found in the precise cross-stream coupling between ventral and dorsal streams (23). In particular, abundant auditory–motor vocal coupling at the prefrontal level might be a prerequisite for volitional control of the vocal apparatus (41, 42, 69).

Several other reasons may exist for the discrepancy between monkeys' ability to learn auditory–motor tasks using their upper extremities and their inability to do so with their vocal apparatus. A first reason may lie in the differing organization of descending motor control. As in humans, movements of the upper limbs (shoulder, elbow, and fingers) in the macaque are under direct, voluntary control from primary motor cortex (48, 70–72), with tactile and proprioceptive feedback operating on the basis of optimal feedback control and state estimation (73). This enables

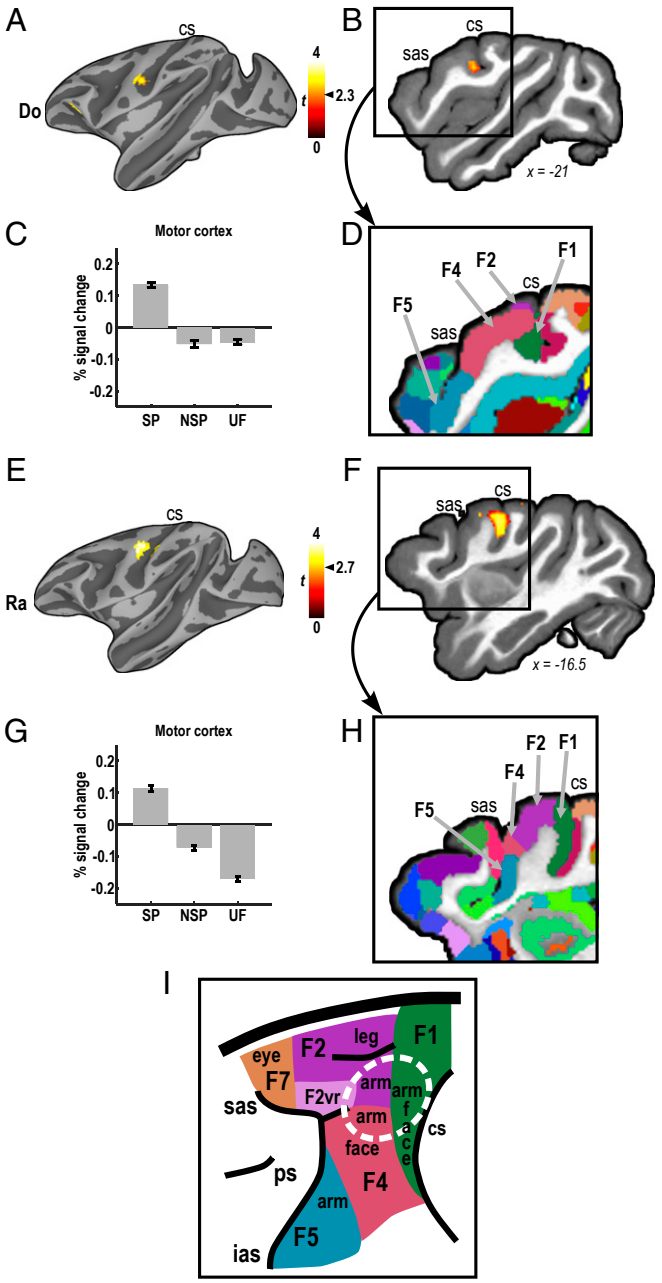

Fig. 3. Activation of motor cortex by listening to self-produced (SP) sound sequences. (*A*) Comparison of BOLD responses evoked by a sound sequence that monkey Do had learned to produce (SP) vs. sound sequences she was passively familiar with (non–self-produced [NSP]) or that were completely unfamiliar (UF) found a peak of activation in the precentral gyrus ($P < 0.01$, uncorrected; $k \geq 10$ voxels). (*B*) The same result shown in volume space. (*C*) Percent signal change evoked by SP, NSP, and UF sequences within the precentral focus defined by significant SP vs. NSP activation; see *SI Appendix*, Fig. S3 for activation by SP vs. NSP and SP vs. UF contrasts. (*D*) The activation mapped to primary motor (F1/M1) and ventral premotor cortex (F4) [D99 atlas segmentation of the same brain (57)]. (*E–H*) Similar results in monkey Ra, including activation of dorsal premotor area F2 ($q < 0.05$, FDR-corrected; $k \geq 25$ voxels; see also *SI Appendix*, Fig. S3). (*I*) Parcellation of the region in the macaque, according to Matelli and Luppino (59). Adjacent arm representations in areas F1, F2, and F4 marked with a white ellipse. F1, F2, F2vr, F4, F5, F7: frontal cortical areas; sulci: cs, central, ias, inferior arcuate, ps, principal, sas, superior arcuate.

monkeys to develop dexterity and tool use with their hands and arms comparable to humans. By contrast, direct descending motor control of the larynx seems to be lacking in the macaque (34). Even if cortical control of the larynx existed in monkeys, however, it would not be sufficient for complex vocal production, because the functions of the larynx are mostly limited to "phonation" (generating voicing and modulating pitch) (74) rather than articulation. Supralaryngeal articulators like tongue, jaw, and lips, however, which are key for human speech production (24), contribute relatively little to the production of monkey calls, despite their general mobility (60, 75). What may be missing, therefore, besides direct descending motor control, is a "command apparatus" (76, 77) orchestrating articulator actions and coordinating them with respiration (24, 78). The general concept of a command apparatus, which was originally proposed by Mountcastle for generating dexterous hand movements (76, 77), can similarly be applied to other cortical areas that also exhibit a high-level "general command function," where output reflects "behavioral goals and not…details of muscular contraction" and encoding suggests an interface between motor and sensory systems. Without such a command apparatus, nonhuman primates may lack an internal model system capable of generating and dynamically controlling motor commands for the production of intended vocal sounds.

In humans, at least three sites may fulfill the requirements for a command apparatus. Damage to (or inactivation of) these regions has been associated with an inability to coordinate complex articulatory movements, termed "apraxia of speech": left ventral sensorimotor cortex (vSMC) (79–82); left precentral gyrus of the insula (83, 84); and Broca's area (specifically, IFG, pars opercularis; BA44) (79, 81). Evidence for sensory input to these motor-related sites is strongest for vSMC (85–88)—where auditory stimuli associated with learned motor behaviors can drive activity in motor cortex, analogous to what we report here

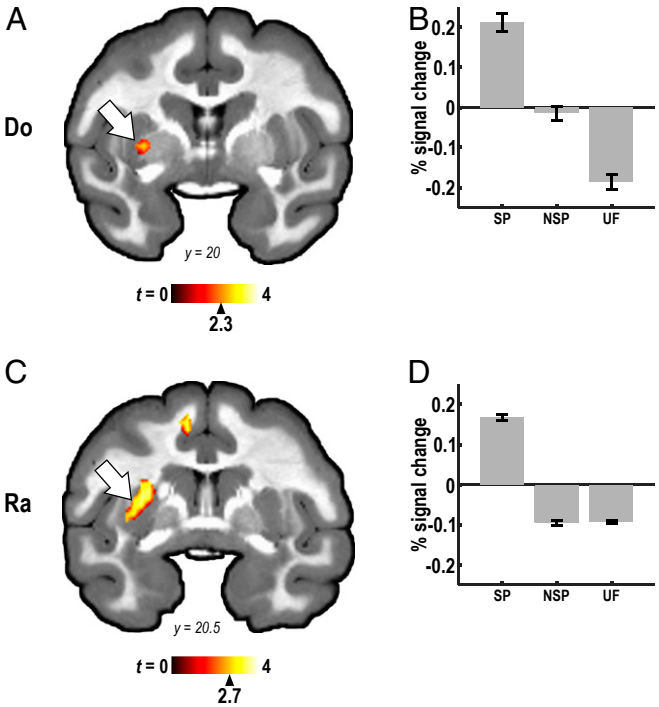

Fig. 4. Putamen activation by listening to SP sequences. (*A*) Comparison of BOLD responses evoked by SP sequence vs. NSP and UF sequences in monkey Do found significant activation in the putamen (arrows; $P < 0.01$, uncorrected; $k \geq 10$ voxels). (*B*) Relative BOLD signal evoked by SP, NSP, and UF sequences in the putamen locus, defined by the SP vs. NSP contrast. (*C* and *D*) Corresponding results for animal Ra ($q < 0.05$, FDR corrected; $k \geq 25$ voxels). See *SI Appendix*, Fig. S5 for location of activation foci revealed by the SP vs., NSP and SP vs. UF contrasts analyzed separately.

for the macaque—and for IFG (27, 45). BA44 may be the most intriguing candidate site as it is often considered the starting point of articulatory planning (89) and has been under marked evolutionary expansion in hominins (60, 90) (see also *SI Appendix*, Supplementary Text). As the endpoint of the auditory dorsal stream, with direct projections from posterior STG (17, 27), BA44 interfaces with high-level auditory structures of the auditory ventral stream (IFG, pars triangularis and orbitalis; BA45). It is in this region, where the major transformation between auditory and motor code might occur: the inverse models that link the intended auditory–sensory outcomes (the vocalization sounds) to motor networks controlling the articulators (41). This transformation (the inverse model) needs to be learned during speech development, in order to be able to produce intended speech sounds. Failure to perform this transformation would thus be another fundamental reason limiting macaques' vocal learning ability.

Inverse models exist in monkeys for the control of the upper extremities [hence enabling reaching and grasping with the hands under visual guidance (91)]. As our study suggests, an inverse model also exists for the production of sounds with the upper extremities, which we demonstrate behaviorally and through activation of motor cortex by the learned sound sequences. By contrast, vocal articulators in the macaque do not seem to be under strict auditory guidance, which would result in a virtual absence of vocal learning required for speech-like abilities (30).

After the discovery of neurons selective for reaching and grasping in premotor area F5 (58, 92), also referred to as "agranular area 6VR" (45) and situated next to BA44 in Broca's region, some have argued for a gestural origin of language (93, 94), whereby spoken expression would have piggybacked on existing mechanisms of visuomotor transformation in frontal and parietal cortex. Given our results, it seems more parsimonious to assume that a more generic, amodal mechanism of sensorimotor control, i.e., internal models, appeared early in evolution, which established the computational principles linking motor with sensory signals. The existence of these mechanisms then may have enabled the parallel development of different modes of expression in language.

The central finding of the present study is that auditory activation of motor-related brain regions (motor cortex and putamen) is selective to sound sequences that the monkeys had learned to self-produce in an auditory–motor task. Learning to play these SP sequences established a specific link between auditory and motor regions that could serve as an internal model (10). Previously known connections from motor to auditory cortical regions ("efference copy") (37, 95) are necessary to inform the sensory system of planned motor articulations that are about to happen, thus anticipating the sensory consequences of action in a forward model. Conversely, connections from auditory to motor regions ("afference copy"), as found in our study, are a basic requirement for the implementation of an inverse model, serving the role of a quick primal sketch of ongoing sensory events (17, 18). This signal is carried by the auditory dorsal stream, a fast, temporally precise pathway connecting posterior auditory areas with premotor cortex (96). The formation of detailed correspondences between auditory and motor patterns can subsequently be established via the auditory ventral stream, a pathway that projects from anterior auditory areas to inferior frontal cortex (17). In all this, learning and predictive coding are essential elements enabling smooth sequential motor behaviors, including articulation of speech or playing a musical instrument.

In addition to cortical structures, and given that speaking is an acquired motor skill, the role of the putamen, as identified here as an important site for auditory–motor learning, should also be considered. The putamen receives an ipsilateral descending projection from primary motor cortex (51) and may play a special role in the above process as a "knitting" device (97) responsible for the coupling of sensory and motor signals into sequences ("chunks") and for learning the conditional probabilities of their co-occurrence (52–54). It is noteworthy, therefore, that molecular analysis of brain regions involved in song and speech of songbirds and humans, respectively (98), has identified the same two regions that were activated in the present study, primary motor cortex and putamen, as showing molecular convergence. They may thus be necessary anatomical and functional components of vocal learning systems. A computational model of birdsong sensorimotor learning provides a specific hypothesis for how BG–forebrain loops could contribute to learning both individual syllables and their proper sequence (99).

In terms of information flow, it would be interesting to confirm the directionality of connections between auditory and motor cortical regions and between motor cortex and putamen with functional connectivity techniques, thus further elucidating their contributions to internal models. Unfortunately, the sparse sampling technique used for auditory functional imaging in our studies, while making it possible to isolate activations to the stimuli from scanner noise, renders it impossible to use directionality estimates such as Granger causality or structural equation modeling, as the temporal evolution of stimulus response is not captured.

Auditory responses in frontal cortex, including motor and premotor regions, have been reported previously in monkey single-unit studies (100, 101), but the precise function of these auditory projections has remained elusive. The present results provide functional evidence for the hypothesized pathway from sensory to motor regions that is critical for learning in current computational models of motor control. Little is known about the anatomical projections from auditory to motor regions. A study by Ward et al. (102) using strychnine-evoked activations suggests auditory input to dorsal and ventral premotor cortex (PMC), and to what was then considered area 4S at the boundary between primary motor cortex and dorsal PMC. The tracer study by Romanski et al. (27) found label in dorsal PMC (area 6d) in addition to dorsolateral prefrontal regions, after injections of area CL in the caudal auditory belt cortex. Whether there are direct projections from auditory regions to M1 is currently unknown. Detailed studies of the anatomical connections between auditory and motor regions are currently under way in our laboratory.

Given these and other reports of a primarily premotor participation in auditory–motor integration [including activation by learned phoneme categories (103), pseudowords (104), and auditory sequences in humans (53, 105)], the predominant activation of primary motor cortex (F1 or M1, depending on nomenclature) in addition to adjacent premotor areas (F2 and F4) (58, 59) came as a surprise. Comparison with electrophysiological mapping studies shows that the activated region in both monkeys contains the arm representations of areas F1, F2, and F4, abutting each other in an apparent overlap (59). This area has recently been given its own identity as a unique functional region (C3) on the basis of resting-state connectivity studies and is involved in shoulder, wrist, and elbow movements (106). The slightly more ventral peak of activation in monkey Do (Fig. 3*A*) extends into area C18 of Goulas et al. (106), which, according to electrophysiological recording studies, contains the macaque "lateral grasping network," a neural substrate for generating purposeful hand actions, including reaching and grasping (107). It appears, therefore, that the entire cortical region in which we found activation by auditory SP sequences is involved in the integration of upper-extremity signals. Its activation by a sound sequence produced with the same effectors indicates the presence of a motor representation of these sounds. Whether and how this region communicates with the parietal command apparatus for sensorimotor integration and control (76, 77) await further studies. Equivalent regions for the control of vocal articulators do not seem to exist in the macaque.

Although our study was primarily designed to understand mechanisms of speech production rather than perception, activation of

primary motor cortex in an auditory–motor task inevitably evokes echoes of "motor theories of speech perception" (108, 109), which often assume a direct connection between auditory and motor areas in the service of speech. Since the time the original theory was formulated, involvement of motor cortex in the processing of receptive speech has been demonstrated with fMRI, transcranial magnetic stimulation, and high-density electrocorticography recordings (85–88, 110). However, as auditory activation of motor areas does not appear to be essential for routine speech recognition, the exact role of this activity has remained elusive and controversial (111). A reinterpretation of these findings in terms of an inverse model may help to clarify the role of these activations.

Our discovery that nonhuman primates can learn to perform a complex auditory–motor task involving both auditory and motor regions suggests that motor cortex activation by sounds is not special to speech. Rather, it is explained by the inverse-model structure of learned auditory–motor behaviors. Just as the evolution of a specialized command apparatus for the neural control of hand and fingers enabled tool use (71, 77), evolution of a special command apparatus for the control of vocal articulators may have enabled speech in primates.

Materials and Methods

Animals. Three adult rhesus monkeys were trained in the behavioral tasks: Ra (male, 8 to 13 y of age over the duration of the study, 12 to 13 kg), Do (female, 7 to 11 y, 5 to 6 kg), and Ch (male, 6 to 9 y, 10 to 12 kg). The animals were surgically implanted with a PEEK headpost (Applied Prototype) secured to the skull with ceramic bone screws (Thomas Recording), plastic strips, and bone cement (Zimmer). Headposts were used to immobilize the head during functional imaging to reduce movement artifacts. Surgical procedures were conducted under aseptic conditions using isoflurane anesthesia, with postoperative analgesic and prophylactic antibiotic treatment. All procedures were approved by the Georgetown University Animal Care and Use Committee and followed the National Research Council's *Guide for the Care and Use of Laboratory Animals* (112).

Behavioral Tasks. Two behavioral tasks were used. In the main task (auditory–motor or monkey piano task) each monkey learned to produce a sequence of eight musical notes by pressing four levers ("piano keys") in a specific order, thus forming an auditory–motor association. A second task [attentional saccade task (113)] was used during fMRI scanning to ensure attention to auditory stimuli and reduce the likelihood of movement. To facilitate parallel learning, the monkey piano task and the attentional task were trained in different locations (home cage vs. sound booth) and used different response methods (lever press vs. saccade) and reward (Fruit Crunchies 190-mg pellets [BioServ] vs. water/diluted fruit juice).

Monkey Piano and SP Sequences. The monkey piano apparatus was designed and built in-house and consisted of an opaque plastic plate, which was temporarily attached to the outside of a monkey home cage for training. Four semitranslucent levers (each of which could be lit with a red light-emitting diode [LED]) and a pellet receptacle were accessible from inside the cage. A pellet delivery device (ENV-203-190IR; Med Associates) and a loudspeaker (TS-G1042R; Pioneer) were placed on the opposite side of the plate, with the loudspeaker delivering sound to the cage via small openings (Fig. 1*A*). The apparatus was controlled by a Power1401 laboratory interface (CED) under control of Spike2 software (CED) and in-house written scripts. Each lever of the apparatus was associated with a musical tone of a fixed pitch in descending order C4, G3, E3, C3 (fundamental frequencies 262, 196, 165, and 131 Hz, respectively, from left to right from the monkey's point of view) (Fig. 1*A*). The tones were digital renderings (44.1 kHz, 16 bit) of single notes generated with VSTi virtual instruments (piano: DSK AkoustiK Keyz, Concert Piano F; trumpet: DSK Brass, Clean Trumpet; cello: DSK Strings, cello; all from DSK Music, https://www.dskmusic.com/). The general frequency range covered by the stimulus overtones was similar between the timbres (*SI Appendix*, Fig. S6). For training, piano tones were used for monkey Ra and trumpet tones for monkeys Do and Ch. After each correct lever press, the associated tone was immediately played by the Power1401 interface via an audio amplifier (AX-496; Yamaha) and the piano loudspeaker. The tone durations were 350 to 500 ms. If another correct lever press occurred while the previous tone was still playing, the previous tone was cut off and the new tone began to play.

Training of the monkeys to produce a sound sequence fell into two phases. In the initial phase (instrumental training, consisting of three steps), the monkeys learned that lever pressing resulted in a sound and, possibly, a reward. In the subsequent phase (sequence training), they learned to press the levers in a predetermined order to produce an auditory–motor sequence.

As the first step of the instrumental training phase, the monkeys were trained to perform an instrumental response of pressing a lever to obtain a reward. All levers were lit with LEDs in the beginning of each trial. If any lever was pressed, all LEDs were turned off, the musical tone associated with the pressed lever was played, and the reward was delivered. Immediately after that, all LEDs were turned back on and the monkey could again press any lever to obtain a reward. In some cases, only a subset of levers was lit and made active (i.e., pressing them would produce a sound and reward) while others were inactive (no light, and no sound or reward when pressed) to prevent the animals from learning to use only particular levers. At the second stage of instrumental training, the monkey was trained to use different levers (in any order) to obtain a reward. The rules were identical as in the first step, except that after reward delivery only levers that had not been pressed were lit and made active, until all four levers were used. In this way, the monkey was forced to use every lever once every four presses on average. In the third and final stage of instrumental training, the monkey learned that obtaining a reward may require several consecutive lever presses. All levers were initially lit. Pressing any lever caused this lever to become inactive (with its LED turned off), while the other, not yet pressed levers remained lit and active. A reward was provided only after all four levers were pressed (in any order), and consequently all LEDs extinguished, followed by offering the choice of all levers again.

After mastering this instrumental training phase, the sequence training phase began, whereby the animals learned to produce a specific auditory–motor sequence (SP sequence). For each monkey, a different SP of eight tones was selected: C4–C4–C3–reward–G3–G3–E3–E3–C3–reward in piano timbre for Ra; E3–C4–G3–reward–C3–C3–E3–G3–C4–reward in trumpet timbre for Do; and C3–E3–G3–reward–C3–E3–C4–reward–G3–C4–reward in trumpet timbre for Ch (Fig. 1*B*). The sequences were created following two rules: each note had to be present exactly two times in the full eight-note sequence, and no identical series of three or more pitches could be present in a passively learned NSP sequence (see below).

Visual guidance with LEDs was initially provided for all presses, but in the course of learning it was gradually reduced down to one to three notes of the sequence (Ra: note 1; Do: notes 1 and 4; Ch: 1, 4, and 7). The goal of the training was to produce temporally consistent performance so that a relatively stable auditory sequence was generated and associated with a sequence of motor actions. For this reason, training progress was judged based on temporal stability and ability to perform over extended periods of time rather than on error rate.

Sound sequences produced by the monkeys naturally acquired an idiosyncratic rhythm imposed by each monkey's habits and by the preceding behavioral context: A repeat of the same note resulted in the shortest interpress interval compared to a note following a different note, and reward delivery caused maximal delay to the following lever press because of the time needed to pick up and eat the pellet (Fig. 1*B*). For this reason, and because of outliers characteristic for a self-paced task (sometimes monkeys took breaks of up to 10 min or more before returning to the task), the performance was examined using median interpress interval calculated for each note in the sequence (Fig. 1 *B* and *C*). Latency distributions in Fig. 1*B* were constructed using 5,000 bins equally spaced on a logarithmic scale in the 0.05- to 10-s range.

For correlation of behavioral performance with fMRI activation (see below), we used a measure of temporal variability of sequence playing, i.e., the interquartile range of interpress intervals, which was normalized (i.e., divided by the median for each sequence position) to account for the effect of behavioral context (see Fig. 1*B*), and averaged across sequence positions.

The self-paced nature of the task led to highly disparate numbers of sequences played per daily session, from (occasionally) as low as 1 to 2, up to ~450 in Ra. Thus, behavioral analyses were conducted in blocks of 50 consecutive sequences (400 notes) played (Fig. 1*C*), with any incomplete sequences at the end of the daily session (as well as rare daily sessions with fewer than 3 completed sequence plays) removed from the analysis. For animals Ra and Ch, timing data were not collected during the initial 26 sessions. In Fig. 1*C*, Ra's data between blocks 550 and 950 are omitted for display purposes.

Ultimately, the stable sound sequences were used as auditory stimuli in fMRI scans (see below). Although we trained three monkeys on the "piano"

task, Ch never achieved a level of performance comparable to the other two animals (Fig. 1*C*), nor was he able to learn the saccade task used in the scanner (see below). Thus, fMRI data were only collected from two monkeys (Ra and Do).

Catch Trials. To test whether the monkeys listened to SP sequences while producing them on the monkey piano, in a subset of sessions for monkeys Ra and Do, lever presses were programmed to produce a different sound than expected, or to produce no sound either in one sequence position, or for the entire eight-note sequence. Such altered sequences were pseudorandomly interspersed between regular-sounding sequences with about 5% frequency. As behavioral effects of an alteration were sometimes delayed, we examined lever press latencies combined for eight consecutive lever presses after the altered/silenced note (or after the first note of the altered/silenced sequence), compared to latencies after unaltered notes or sequences. Because different behavioral contexts led to vastly different press latencies depending on sequence position (see Fig. 1*B*), raw latencies were first log-transformed and then normalized separately for each sequence position. For that purpose, the distribution's mean was subtracted and the result divided by its SD. Only then were catch-trial data separated from unaltered-trial data for comparison across sequence positions. Fig. 1*D* shows the effect of a single note alteration: Note C4 in piano timbre was played in sequence position 7 for monkey Ra (original note: E3 in piano timbre) and in sequence position 1 for monkey Do (original note: E3 in trumpet timbre). *SI Appendix*, Fig. S2 shows the effect of silencing the entire sequence.

NSP Sequences. The goal of the study was to identify brain regions specifically activated by sound sequences associated with motor actions. To control for a possible confound of familiarity, the animals were passively exposed to NSP sequences, which were similar to SP in structure and familiarity, but were not associated with any motor actions. Like SP, NSP consisted of eight notes of pitches C3, E3, G3, and C4, each pitch repeating twice within the sequence. They differed from SP in the order of notes and, to accentuate the difference, in timbre.

One potentially confounding difference between SP and NSP sequences was that SP, but not NSP sequences, were associated not only with motor actions but also with reward. However, we decided not to provide rewards during passive exposure to NSP to avoid associating with NSP any motor actions that resulted simply from picking up or eating the reward.

The NSP sequences were as follows: C3–E3–G3–pause–C3–E3–C4–pause–G3–C4–pause in trumpet timbre for Ra, and C4–C4–C3–pause–G3–G3–E3–E3–C3–pause in piano timbre for Do and Ch. Note that Ra's SP is identical to Do and Ch's NSP, and Ch's SP is identical to Ra's NSP. This symmetric design was intended to allow for reciprocal playback to ensure high familiarity of NSP. The training of the monkeys was initially conducted in pairs in one room, so that when one monkey played the monkey piano learning his or her SP, another monkey in the room was passively exposed to and familiarized with the same melody, which became NSP for that second monkey. Subsequently, exposure to NSP was largely realized via playback of recordings in the holding room.

UF Sequences. UF sequences were generated automatically with a Matlab script, using the following rules: eight notes; same four pitches as SP and NSP; timbre, cello; in all eight positions, the pitch had to be different from the pitch in the same position in the same monkey's SP and NSP sequences. In addition, the temporal patterning of SP and NSP was simulated according to the behavioral context (see above and Fig. 1*B*). Namely, a note following the same note was preceded by an internote interval of nominally 0.15 s (onset-to-onset), a note following another note was preceded by nominally 0.3 s, and a note following simulated reward was preceded by nominally 1 s. Eighteen UF sequences were generated for Ra and 15 for Do, and each of those was generated in three variants, with actual internote intervals randomly deviating from the nominal values by up to ±25% (uniform distribution).

Saccade Task for fMRI. In order to assure a constant attention level during BOLD imaging, monkeys performed an attentional saccade task, as previously described (113). The task was a go–no-go auditory detection task, white noise being the target stimulus.

Initial training in this task was performed outside the MRI scanner. The animal was lying in sphinx position and head-fixed in an MRI-compatible primate chair (Applied Prototype) or an in-house built primate chair of similar size and orientation, placed inside a sound booth simulating the scanner environment. In some sessions, simulated MRI scanner noise (collected from various scanner noise recordings available on YouTube) was played via a loudspeaker (MSP3; Yamaha) to acclimate the animals with the conditions inside the scanner.

Auditory stimuli, both in training and during fMRI acquisition, were delivered via modified electrostatic in-ear headphones (SRS-005S + SRM-252S; STAX) mounted on ear-mold impressions of the monkey pinna (Starkey) and covered with a custom-made earmuff system for scanner noise attenuation. An LCD monitor was used to present a central red fixation spot. Eye movements were monitored using an infrared eye-tracking system (ETL-200; ISCAN) with the analog output sampled with an analog-to-digital converter (USB-6218; National Instruments). The task (including triggering of fMRI acquisition [see below]) was controlled with Presentation software (Neurobehavioral Systems) and custom-made scripts.

In the task, the monkey initiated a trial by holding fixation on a central red spot. Next, a block of auditory stimuli was presented while fixation was held. Breaking fixation stopped auditory presentation and restarted the trial. For each session, a minimum duration of auditory stimulation was set, and for each trial the stimuli were randomly selected from a predefined list and arranged serially into a stimulus block with an interstimulus interval of 0.2 to 1.5 s (depending on the stage of training; the interval was often jittered) so that the total duration of the stimulus block was equal to or larger than the minimum duration. The minimum duration was initially short (e.g., 1 to 2 s), and then, as training progressed, it was gradually increased to 8.7 to 9.7 s. After presentation of the stimulus block, the target WN burst was played, and the monkey had another second to perform a saccade to the left, which was rewarded with water/juice and with a yellow confirmation spot occurring in the expected saccade target location. In about half of the trials, silence was used instead of auditory stimuli, and the monkey had to keep fixation until the target sound was played. Such silent trials were then used as baseline in the fMRI experiment.

Auditory stimuli used during training were tones of various frequencies, bandpass noise bursts of various center frequencies and bandwidths, primate vocalizations, and environmental sounds. Neither SP, NSP, UF, nor any similar sequences or individual instrumental sounds were used for training of the saccade task; the monkeys were exposed to stimuli of this type only when playing SP on the monkey piano, or while being exposed passively to NSP, or during fMRI scans (see below). All stimuli (including SP, NSP, and UF used in scanning) were loudness-matched to equal maximum RMS amplitude in a 200-ms sliding window, taking into account frequency-dependent sensitivity of monkey hearing (114) [similar to the dB(A) scale used for humans; see also ref. 115], and frequency response of the presentation system (obtained with a probe microphone [Brüel & Kjær; type 4182 SPL meter] inserted in the ear mold of an anesthetized monkey). The stimuli were amplified (RA-300, Alesis, or SLA-1, A.R.T.) and delivered at an intensity equivalent to ~80-dB SPL at 1 kHz.

fMRI Scanning. We used a behavior-driven sparse sampling paradigm, acquiring a single volume per trial with a delay that matches the predicted peak of the evoked hemodynamic response (116). This prevents contamination of auditory stimuli with gradient noise, and consequently, contamination of the stimulus-evoked BOLD response with the scanner-noise-evoked response.

Monkeys were placed in a 3-T Magnetom Tim Trio (Siemens) 60-cm horizontal-bore MRI scanner lying in sphinx position in an MRI-compatible chair (Applied Prototype), with their head fixed to the chair structure using the implanted headpost. A 12- or 10-cm saddle-shaped radiofrequency coil (Windmiller Kolster Scientific) was placed over the head and covered the entire brain. The time series consisted of gradient-echo echo-planar (GE-EPI) whole-brain images obtained in a sparse acquisition design. During signal acquisition, monkeys were working in the saccade task as described above, with the following differences: The visual fixation and confirmation spots were backprojected onto a semitranslucent screen; the required accuracy of fixation and saccades was somewhat relaxed to account for the animal's generally being more prone to distraction in the scanner, and for the inability to calibrate the system as accurately as in the sound booth due to the required safety distance of the eye tracker from the scanner bore. The stimuli used were SP and NSP, and (except for some early scans; see below) UF, each in several variants with slightly differing presentation tempo to account for idiosyncratic differences in performance by the monkeys. The interstimulus interval was 0.5 to 1 s, and the minimum duration of the stimulus block was at least 9.2 s. This resulted in two to three presentations of a sequence per trial. Seven seconds after the stimulus block started, an fMRI volume acquisition was triggered. Thus, with an acquisition time (TA) of 2.18 s, acquisition was completed before WN presentation and saccade. If the monkey broke fixation before the acquisition trigger, the trial was restarted. If fixation was broken after the trigger, the trial continued but was censored from statistical analyses.

The order of trials was as follows: 6× SP, 3× silence, 6× NSP, 3× silence, 6× UF, 3× silence, typically 10 cycles, i.e., 270 trials (and volumes) per scanning session. In early scans not involving UF, the order was as follows: SP, silence, NSP, silence, typically 40 cycles, 160 trials (and volumes) per session.

Individual volumes with 23 (Do) or 25 (Ra) ordinal slices were acquired with an interleaved single-shot GE-EPI sequence (echo time [TE], 34 ms; TA, 2.18 s; flip angle, 90°; field of view [FOV], 100 × 100 mm^2; matrix size, 66 × 66 voxels; slice thickness, 1.9 mm; voxel size, 1.5 × 1.5 × 1.9 mm^3). For overlaying the functional images, high-resolution structural images were acquired in a separate session under general isoflurane anesthesia (magnetization-prepared rapid gradient-echo sequence; voxel size, 0.5 × 0.5 × 0.5 mm^3; four to five averages; TE, 3.0 ms; repetition time [TR], 2.5 s; flip angle, 8°; FOV, 116 × 96 × 128 mm^3; matrix, 232 × 192 × 256 voxels).

After censoring (see below), 12 scanning sessions from each Ra and Do were used for the analysis. Of those, the first four Do and three Ra sessions did not involve UF. In addition, Ra was not required to perform the saccade task during these first three sessions, but the visual and auditory stimuli and reward were delivered as if he performed the task perfectly.

fMRI Data Analyses. Data were analyzed with AFNI (Scientific and Statistical Computing Core [SSCC], National Institute of Mental Health [NIMH]). T2*-weighted echo-planar images were coarsely aligned across sessions with a full affine transformation. Volumes were then individually aligned, with a bilinear warp (*3dWarpDrive*), to the median image of a representative session. Data were smoothed with a three-dimensional Gaussian kernel (3-mm full width at half-maximum) and normalized by the median. To detect volumes with excessive motion artifact, the fraction of outliers in each volume was assessed. Voxel intensities were deemed outliers if they exceeded:

$$\mathrm{MAD} * \Phi^{-1}\left(\frac{1-\alpha}{n}\right)\sqrt{\frac{\pi}{2}}, \alpha = 10^{-3},$$

where MAD is the median absolute deviation, $\Phi^{-1}$is the inverse Gaussian CDF, n is the number of temporal observations, and α is the threshold parameter. Volumes whose outlier fraction exceeded 0.0005 were censored from analyses. For anatomical localization of results, the D99 atlas template (57) was aligned to a high-resolution T1 image with a nonlinear warp (full affine transform followed by incremental polynomial warps, *3dQwarp*). The T1 image was then similarly aligned to the median EPI, and that alignment was applied to D99 as well.

To assess data quality within session, each session's data were fit with ordinary least squares to a regression model containing terms for each stimulus condition (SP, NSP, and UF) and 20 nuisance terms: 12 for head motion (3 for displacement, 3 for rotation, and 6 for their temporal derivatives) and 8 Lagrange polynomial terms serving as variable high-pass filters to model baseline fluctuation (sampled according to the acquisition times for each volume, which were triggered by the monkey). A general linear test for all stimulus conditions vs. baseline (i.e., null trials) was performed to assess activation in auditory regions (similar to Fig. 2 but individually for each session). Sessions where auditory cortex activation was not observed (one session each, Do and Ra) were censored from further analyses. Additionally, three sessions were censored from analyses for poor alignment to template (one session, Do), and excessive motion artifact (one session each, Do and Ra; assessed via cine loop).

Sessions were concatenated to construct multisession regression models, one per monkey, yielding 1,258 (Ra) and 1,140 (Do) volumes per model. Separate baseline fluctuation terms were included for each session in multisession models. Head motion terms were concatenated across sessions. A mean-centered behavioral variability term (based on interquartile range of interpress intervals, as described above) was calculated for the out-of-scanner block of monkey piano training encompassing each scanning session (cf. Fig. 1*C*) and included in the model. This parameter varied as a function of session and was specified only for volumes in the SP condition. Apart from the performance term result (i.e., *SI Appendix*, Fig. S4), which reflects a test of a single regression coefficient, all reported single-animal effects are for either general linear tests (i.e., Fig. 2) or contrast tests (i.e., Figs. 3 and 4 and *SI Appendix*, Figs. S3, S5, and S7) over sets of regression coefficients. To be directionally consistent with the main hypothesis (i.e., that audio-motor learning induces the formation of neuronal assemblies in frontal cortex that are specifically responsive to motor-associated auditory inputs), the contrast tests (i.e., Figs. 3 and 4 and *SI Appendix*, Figs. S3, S5, and S7) were one-tailed. To identify stimulus processing in the ascending auditory pathway, one-tailed tests were used for Fig. 2.

For additional results visualization, cortical surface models were constructed from high-resolution T1 images using FreeSurfer (Martinos Center, Massachusetts General Hospital). Statistical maps were then projected onto the cortical sheet with SUMA (SSCC, NIMH) (Fig. 3 *A* and *E*).

Given the small sample size ($n = 2$), group-level effects were assessed via the anatomical concordance of effects observed at the single-animal level (i.e., activation foci that mapped to the same anatomical structure or cortical region were considered to have replicated across animals). For supplemental analyses of the performance term (*SI Appendix*, Fig. S4), the activation reported is concordant with the main findings (Figs. 3 and 4 and *SI Appendix*, Fig. S3).

EMG and Movement Recordings. To assess whether the increased activation of the motor cortex by listening to SP sequences compared to NSP and UF sequences (Fig. 3) could be related to a higher level of motor activity driven by SP sequences (rather than being auditory-driven), we measured arm movements and electrical activity of shoulder muscles while monkey Ra was listening to the sequences and to control stimuli. After shaving hair from the left arm and shoulder, two adhesive surface electrodes (Red Dot 2560; 3M) were placed on the skin over the deltoid muscle for EMG recording: common ground and negative electrode at the distal end, positive electrode proximally over the muscle, while the monkey was head-fixed in a monkey chair in a sound-isolated booth. Electrode gel (Gel 101; Biopac) was used to improve electrode contact, resulting in a 10- to 40-kΩ impedance. A one-axis accelerometer (MMA1250KEG; NXP Semiconductors) was embedded in epoxy resin together with a supporting circuit and affixed to the arm just above the elbow using another (inactive) 2560 electrode as a sticky pad, with the accelerometer axis oriented approximately perpendicular to the humerus and antero-posterior when the arm was lowered along the body. EMG activity was amplified, filtered (1 Hz to 5 kHz, 60-Hz notch; model 1902; CED), and sampled along with the accelerometer output (both at 20 kHz) by a Power 1401mkII interface, using Spike2 software and custom-made scripts, which also presented auditory stimuli via an attenuator (model 3505; CED) and a loudspeaker (MSP3; Yamaha) at approximately 65 to 70 dB(A) (except silence). The stimuli were the same SP, NSP, and UF melodies that were used during fMRI scanning (the latter chosen randomly in each block from the set of UF stimuli, as described above); two acoustic controls that were not sequences (a series of environmental sounds [ES] recorded in the monkey facility and a series of six 300-ms WN bursts [WN]); and silence (Sil). The stimulus durations were as follows: SP, 2.3 s; NSP, 5.5 s; UF, 3.3 to 4.1 s; ES, 5.7 s; WN, 5.3 s; and Sil, 6 s. A block of six stimuli was played in random order with 3.5- to 4.5-s interstimulus intervals for 20 block repeats, with between-block intervals of 3.5 s or more. The monkey was occasionally rewarded with a treat given directly to the mouth during between-block intervals. Four recording sessions were conducted, two of which included arm movement recording with the accelerometer.

The digitized EMG signal was additionally filtered (notch filters at 60, 180, and 300 Hz, and bandpass 5 to 200 Hz). DC offset was removed from the digitized accelerometer signal by local subtraction of the mean in a 0.1-s sliding window to eliminate the effect of sustained arm position relative to Earth's gravity. Finally, both signals were RMS-averaged with a 0.1-s sliding window.

Two measures were compared between the presentation of the SP sequence and of the other stimuli for both movement and EMG: mean RMS-averaged signal, and the fraction of time during which the RMS-averaged signal remained above a threshold established visually as a constant level that remained above baseline/noise but was typically crossed by activity bursts (*SI Appendix*, Fig. S1*E*).

Analysis was performed over the time period from stimulus onset to 3 s beyond stimulus end, as well as separately for the stimulus period and for the 3 s after the stimulus, in order to be able to detect any motor activity evoked by SP under different hypothetical scenarios, like SP immediately causing movement activity, or activity generated to continue playing the sequence after it has ended. Despite purposefully applying lenient statistical criteria (uncorrected multiple-comparison tests, working against the hypothesis), neither approach detected significantly increased EMG activity or overt movement when the monkey listened to SP compared to other melodies or control stimuli (*SI Appendix*, Fig. S1). Only in one case one measure of EMG activity was higher during SP than during silence (*SI Appendix*, Fig. S1*D*). Recording quality was confirmed by cross-correlating the accelerometer and EMG signals; as expected, the motor output lagged behind electrical activity (*SI Appendix*, Fig. S1*F*).

Data and Materials Availability. MRI data associated with the figures in this work have been deposited on PRIME-DE (http://fcon_1000.projects.nitrc.org/indi/PRIME/georgetown.html).

ACKNOWLEDGMENTS. We thank Drs. Max Riesenhuber, Peter Turkeltaub, and Xiong Jiang, and Ms. Jessica Jacobs for comments on the manuscript; Dr. Lars Rogenmoser for advice on EMG recordings; and Jeff Bloch for help with data collection and analysis. This research was supported by NIH Grant R01DC014989 (J.P.R.) and a Partnerships for International Research and Education (PIRE) grant from the National Science Foundation (PIRE-OISE-0730255). I.P.J. and M.S. were additionally funded by the Academy of Finland (Grant 276643).

1. W. J. M. Levelt, *Speaking: From Intention to Articulation* (MIT Press, 1989).
2. R. Cusack, C. J. Wild, L. Zubiaurre-Elorza, A. C. Linke, Why does language not emerge until the second year? *Hear. Res.* **366**, 75–81 (2018).
3. M. I. Jordan, D. E. Rumelhart, Forward models: Supervised learning with a distal teacher. *Cogn. Sci.* **16**, 307–354 (1992).
4. C. R. Smith, Residual hearing and speech production in deaf children. *J. Speech Hear. Res.* **18**, 795–811 (1975).
5. E. von Holst, H. Mittelstaedt, Das Reafferenzprinzip–Wechselwirkungen zwischen Zentralnervensystem und Peripherie. *Naturwissenschaften* **37**, 464–476 (1950).
6. W. Hershberger, Afference copy, the closed-loop analogue of von Holst's efference copy. *Cybern. Forum* **8**, 97–102 (1976).
7. W. T. Powers, The reafference principle and control theory. www.livingcontrolsystems.com/intro_papers/Reafference_principle.pdf. Accessed 28 April 2020.
8. D. M. Wolpert, Z. Ghahramani, M. I. Jordan, An internal model for sensorimotor integration. *Science* **269**, 1880–1882 (1995).
9. M. Kawato, Internal models for motor control and trajectory planning. *Curr. Opin. Neurobiol.* **9**, 718–727 (1999).
10. D. M. Wolpert, Z. Ghahramani, Computational principles of movement neuroscience. *Nat. Neurosci.* **3** (suppl.), 1212–1217 (2000).
11. G. Hickok, J. Houde, F. Rong, Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron* **69**, 407–422 (2011).
12. F. H. Guenther, *Neural Control of Speech* (MIT Press, 2016).
13. F. H. Guenther, A neural network model of speech acquisition and motor equivalent speech production. *Biol. Cybern.* **72**, 43–53 (1994).
14. J. A. Tourville, K. J. Reilly, F. H. Guenther, Neural mechanisms underlying auditory feedback control of speech. *Neuroimage* **39**, 1429–1443 (2008).
15. J. A. Tourville, F. H. Guenther, The DIVA model: A neural theory of speech acquisition and production. *Lang. Cogn. Process.* **26**, 952–981 (2011).
16. E. Todorov, Optimality principles in sensorimotor control. *Nat. Neurosci.* **7**, 907–915 (2004).
17. J. P. Rauschecker, S. K. Scott, Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nat. Neurosci.* **12**, 718–724 (2009).
18. J. P. Rauschecker, An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hear. Res.* **271**, 16–25 (2011).
19. J. F. Houde, S. S. Nagarajan, Speech production as state feedback control. *Front. Hum. Neurosci.* **5**, 82 (2011).
20. G. Hickok, Computational neuroanatomy of speech production. *Nat. Rev. Neurosci.* **13**, 135–145 (2012).
21. J. F. Houde, E. F. Chang, The cortical computations underlying feedback control in vocal production. *Curr. Opin. Neurobiol.* **33**, 174–181 (2015).
22. B. Parrell, V. Ramanarayanan, S. Nagarajan, J. Houde, The FACTS model of speech motor control: Fusing state estimation and task-based control. *PLoS Comput. Biol.* **15**, e1007321 (2019).
23. I. Bornkessel-Schlesewsky, M. Schlesewsky, S. L. Small, J. P. Rauschecker, Neurobiological roots of language in primate audition: Common computational properties. *Trends Cogn. Sci.* **19**, 142–150 (2015).
24. J. P. Rauschecker, Where did language come from? Precursor mechanisms in nonhuman primates. *Curr. Opin. Behav. Sci.* **21**, 195–204 (2018).
25. R. J. Zatorre, J. L. Chen, V. B. Penhune, When the brain plays music: Auditory-motor interactions in music perception and production. *Nat. Rev. Neurosci.* **8**, 547–558 (2007).
26. I. Wollman, V. Penhune, M. Segado, T. Carpentier, R. J. Zatorre, Neural network retuning and neural predictors of learning success associated with cello training. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E6056–E6064 (2018).
27. L. M. Romanski *et al*., Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat. Neurosci.* **2**, 1131–1136 (1999).
28. R. M. Seyfarth, D. L. Cheney, P. Marler, Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science* **210**, 801–803 (1980).
29. A. A. Ghazanfar, M. D. Hauser, The neuroethology of primate vocal communication: Substrates for the evolution of speech. *Trends Cogn. Sci.* **3**, 377–384 (1999).
30. C. I. Petkov, E. D. Jarvis, Birds, primates, and spoken language origins: Behavioral phenotypes and neurobiological substrates. *Front. Evol. Neurosci.* **4**, 12 (2012).
31. G. Coudé *et al*., Neurons controlling voluntary vocalization in the macaque ventral premotor cortex. *PLoS One* **6**, e26822 (2011).
32. S. R. Hage, A. Nieder, Single neurons in monkey prefrontal cortex encode volitional initiation of vocalizations. *Nat. Commun.* **4**, 2409 (2013).
33. U. Jürgens, Neural pathways underlying vocal control. *Neurosci. Biobehav. Rev.* **26**, 235–258 (2002).
34. K. Simonyan, U. Jürgens, Efferent subcortical projections of the laryngeal motorcortex in the rhesus monkey. *Brain Res.* **974**, 43–59 (2003).
35. J. C. Dunn, J. B. Smaers, Neural correlates of vocal repertoire in primates. *Front. Neurosci.* **12**, 534 (2018).
36. P. Müller-Preuss, D. Ploog, Inhibition of auditory cortical neurons during phonation. *Brain Res.* **215**, 61–76 (1981).
37. S. J. Eliades, X. Wang, Dynamics of auditory-vocal interaction in monkey auditory cortex. *Cereb. Cortex* **15**, 1510–1523 (2005).
38. S. P. Ó Scalaidhe, F. A. W. Wilson, P. S. Goldman-Rakic, Areal segregation of face-processing neurons in prefrontal cortex. *Science* **278**, 1135–1138 (1997).
39. J. Fuster, *Cortex and Mind: Unifying Cognition* (Oxford University Press, 2003).
40. L. M. Romanski, B. B. Averbeck, M. Diltz, Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. *J. Neurophysiol.* **93**, 734–747 (2005).
41. M. Petrides, G. Cadoret, S. Mackey, Orofacial somatomotor responses in the macaque monkey homologue of Broca's area. *Nature* **435**, 1235–1238 (2005).
42. S. R. Hage, A. Nieder, Dual neural network model for the evolution of speech and language. *Trends Neurosci.* **39**, 813–829 (2016).
43. K. Amunts, K. Zilles, Architecture and organizational principles of Broca's region. *Trends Cogn. Sci.* **16**, 418–426 (2012).
44. M. Petrides, D. N. Pandya, Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biol.* **7**, e1000170 (2009).
45. S. Frey, S. Mackey, M. Petrides, Cortico-cortical connections of areas 44 and 45B in the macaque monkey. *Brain Lang.* **131**, 36–55 (2014).
46. A. D. Friederici, Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci.* **6**, 78–84 (2002).
47. E. Fedorenko, I. A. Blank, Broca's area is not a natural kind. *Trends Cogn. Sci.* **24**, 270–284 (2020).
48. R. N. Lemon, Descending pathways in motor control. *Annu. Rev. Neurosci.* **31**, 195–218 (2008).
49. T. Proffitt *et al*., Analysis of wild macaque stone tools used to crack oil palm nuts. *R. Soc. Open Sci.* **5**, 171904 (2018).
50. S. Miyachi, O. Hikosaka, K. Miyashita, Z. Kárádi, M. K. Rand, Differential roles of monkey striatum in learning of sequential hand movement. *Exp. Brain Res.* **115**, 1–5 (1997).
51. R. S. Turner, M. R. DeLong, Corticostriatal activity in primary motor cortex of the macaque. *J. Neurosci.* **20**, 7096–7108 (2000).
52. Y. Ueda, M. Kimura, Encoding of direction and combination of movements by primate putamen neurons. *Eur. J. Neurosci.* **18**, 980–994 (2003).
53. A. M. Leaver, J. Van Lare, B. Zielinski, A. R. Halpern, J. P. Rauschecker, Brain activation during anticipation of sound sequences. *J. Neurosci.* **29**, 2477–2485 (2009).
54. S. Dehaene, F. Meyniel, C. Wacongne, L. Wang, C. Pallier, The neural representation of sequences: From transition probabilities to algebraic patterns and linguistic trees. *Neuron* **88**, 2–19 (2015).
55. J. H. Kaas, T. A. Hackett, Subdivisions of auditory cortex and processing streams in primates. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11793–11799 (2000).
56. J. P. Rauschecker, B. Tian, Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11800–11806 (2000).
57. C. Reveley *et al*., Three-dimensional digital template atlas of the macaque brain. *Cereb. Cortex* **27**, 4463–4477 (2017).
58. M. Matelli, G. Luppino, G. Rizzolatti, Patterns of cytochrome oxidase activity in the frontal agranular cortex of the macaque monkey. *Behav. Brain Res.* **18**, 125–136 (1985).
59. M. Matelli, G. Luppino, Parietofrontal circuits for action and space perception in the macaque monkey. *Neuroimage* **14**, S27–S32 (2001).
60. W. T. Fitch, The biology and evolution of speech: A comparative analysis. *Annu. Rev. Linguist.* **4**, 255–279 (2018).
61. G. Hickok, D. Poeppel, Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition* **92**, 67–99 (2004).
62. D. Poeppel, K. Emmorey, G. Hickok, L. Pylkkänen, Towards a new neurobiology of language. *J. Neurosci.* **32**, 14125–14131 (2012).
63. J. P. Rauschecker, Cortical processing of complex sounds. *Curr. Opin. Neurobiol.* **8**, 516–521 (1998).
64. I. DeWitt, J. P. Rauschecker, Phoneme and word recognition in the auditory ventral stream. *Proc. Natl. Acad. Sci. U.S.A.* **109**, E505–E514 (2012).
65. I. DeWitt, J. P. Rauschecker, Wernicke's area revisited: Parallel streams and word processing. *Brain Lang.* **127**, 181–191 (2013).
66. J. K. Rilling, Comparative primate neurobiology and the evolution of brain language systems. *Curr. Opin. Neurobiol.* **28**, 10–14 (2014).
67. F. Balezeau *et al*., Primate auditory prototype in the evolution of the arcuate fasciculus. *Nat. Neurosci.* **23**, 611–614 (2020).
68. I. Bornkessel-Schlesewsky, M. Schlesewsky, S. L. Small, J. P. Rauschecker, Response to Skeide and Friederici: The myth of the uniquely human "direct" dorsal pathway. *Trends Cogn. Sci.* **19**, 484–485 (2015).
69. S. R. Hage, A. Nieder, Audio-vocal interaction in single neurons of the monkey ventrolateral prefrontal cortex. *J. Neurosci.* **35**, 7030–7040 (2015).
70. M. Graziano, The organization of behavioral repertoire in motor cortex. *Annu. Rev. Neurosci.* **29**, 105–134 (2006).
71. J.-A. Rathelot, P. L. Strick, Subdivisions of primary motor cortex based on cortico-motoneuronal cells. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 918–923 (2009).
72. K. V. Shenoy, M. Sahani, M. M. Churchland, Cortical control of arm movements: A dynamical systems perspective. *Annu. Rev. Neurosci.* **36**, 337–359 (2013).
73. R. Shadmehr, J. W. Krakauer, A computational neuroanatomy for motor control. *Exp. Brain Res.* **185**, 359–381 (2008).
74. B. K. Dichter, J. D. Breshears, M. K. Leonard, E. F. Chang, The control of vocal pitch in human laryngeal motor cortex. *Cell* **174**, 21–31.e9 (2018).

75. H. Ackermann, S. R. Hage, W. Ziegler, Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behav. Brain Sci.* **37**, 529–546 (2014).
76. V. B. Mountcastle, J. C. Lynch, A. Georgopoulos, H. Sakata, C. Acuna, Posterior parietal association cortex of the monkey: Command functions for operations within extrapersonal space. *J. Neurophysiol.* **38**, 871–908 (1975).
77. J.-A. Rathelot, R. P. Dum, P. L. Strick, Posterior parietal cortex contains a command apparatus for hand movements. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 4255–4260 (2017).
78. M. Belyk, S. Brown, The origins of the vocal brain in humans. *Neurosci. Biobehav. Rev.* **77**, 177–193 (2017).
79. A. E. Hillis *et al*., Re-examining the brain regions crucial for orchestrating speech articulation. *Brain* **127**, 1479–1487 (2004).
80. A. Basilakos, C. Rorden, L. Bonilha, D. Moser, J. Fridriksson, Patterns of poststroke brain damage that predict speech production errors in apraxia of speech and aphasia dissociate. *Stroke* **46**, 1561–1566 (2015).
81. M. A. Long *et al*., Functional segregation of cortical regions underlying speech timing and articulation. *Neuron* **89**, 1187–1193 (2016).
82. A. Basilakos, K. G. Smith, P. Fillmore, J. Fridriksson, E. Fedorenko, Functional characterization of the human speech articulation network. *Cereb. Cortex* **28**, 1816–1830 (2018).
83. N. F. Dronkers, A new brain region for coordinating speech articulation. *Nature* **384**, 159–161 (1996).
84. K. Chenausky, S. Paquette, A. Norton, G. Schlaug, Apraxia of speech involves lesions of dorsal arcuate fasciculus and insula in patients with aphasia. *Neurol. Clin. Pract.* **10**, 162–169 (2020).
85. S. M. Wilson, A. P. Saygin, M. I. Sereno, M. Iacoboni, Listening to speech activates motor areas involved in speech production. *Nat. Neurosci.* **7**, 701–702 (2004).
86. J. I. Skipper, H. C. Nusbaum, S. L. Small, Listening to talking faces: Motor cortical activation during speech perception. *Neuroimage* **25**, 76–89 (2005).
87. F. Pulvermüller *et al*., Motor cortex maps articulatory features of speech sounds. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 7865–7870 (2006).
88. C. Cheung, L. S. Hamiton, K. Johnson, E. F. Chang, The auditory representation of speech sounds in human motor cortex. *eLife* **5**, e12577 (2016).
89. R. J. S. Wise, Language systems in normal and aphasic human subjects: Functional imaging studies and inferences from animal studies. *Br. Med. Bull.* **65**, 95–119 (2003).
90. N. M. Schenker *et al*., Broca's area homologue in chimpanzees (*Pan troglodytes*): Probabilistic mapping, asymmetry, and comparison to humans. *Cereb. Cortex* **20**, 730–742 (2010).
91. M. Jeannerod, M. A. Arbib, G. Rizzolatti, H. Sakata, Grasping objects: The cortical mechanisms of visuomotor transformation. *Trends Neurosci.* **18**, 314–320 (1995).
92. V. Gallese, L. Fadiga, L. Fogassi, G. Rizzolatti, Action recognition in the premotor cortex. *Brain* **119**, 593–609 (1996).
93. G. Rizzolatti, M. A. Arbib, Language within our grasp. *Trends Neurosci.* **21**, 188–194 (1998).
94. M. C. Corballis, The gestural origins of language. *Wiley Interdiscip. Rev. Cogn. Sci.* **1**, 2–7 (2010).
95. J. F. Houde, S. S. Nagarajan, K. Sekihara, M. M. Merzenich, Modulation of the auditory cortex during speech: An MEG study. *J. Cogn. Neurosci.* **14**, 1125–1138 (2002).
96. P. Kuśmierek, J. P. Rauschecker, Selectivity for space and time in early areas of the auditory dorsal stream in the rhesus monkey. *J. Neurophysiol.* **111**, 1671–1685 (2014).
97. J. P. Rauschecker, Is there a tape recorder in your head? How the brain stores and retrieves musical melodies. *Front. Syst. Neurosci.* **8**, 149 (2014).
98. A. R. Pfenning *et al*., Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science* **346**, 1256846 (2014).
99. T. W. Troyer, A. J. Doupe, An associational model of birdsong sensorimotor learning I. Efference copy and the learning of song syllables. *J. Neurophysiol.* **84**, 1204–1223 (2000).
100. M. S. A. Graziano, L. A. J. Reiss, C. G. Gross, A neuronal representation of the location of nearby sounds. *Nature* **397**, 428–430 (1999).
101. C. T. Miller, A. W. Thomas, S. U. Nummela, L. A. de la Mothe, Responses of primate frontal cortex neurons during natural vocal communication. *J. Neurophysiol.* **114**, 1158–1171 (2015).
102. A. A. Ward Jr, J. K. Peden, O. Sugar, Cortico-cortical connections in the monkey with special reference to area 6. *J. Neurophysiol.* **9**, 453–461 (1946).
103. M. A. Chevillet, X. Jiang, J. P. Rauschecker, M. Riesenhuber, Automatic phoneme category selectivity in the dorsal auditory stream. *J. Neurosci.* **33**, 5208–5215 (2013).
104. A. M. Rauschecker, A. Pringle, K. E. Watkins, Changes in neural activity associated with learning to articulate novel auditory pseudowords by covert repetition. *Hum. Brain Mapp.* **29**, 1231–1242 (2008).
105. J. L. Chen, C. Rae, K. E. Watkins, Learning to play a melody: An fMRI study examining the formation of auditory-motor associations. *Neuroimage* **59**, 1200–1208 (2012).
106. A. Goulas *et al*., Intrinsic functional architecture of the macaque dorsal and ventral lateral frontal cortex. *J. Neurophysiol.* **117**, 1084–1099 (2017).
107. E. Borra, M. Gerbella, S. Rozzi, G. Luppino, The macaque lateral grasping network: A neural substrate for generating purposeful hand actions. *Neurosci. Biobehav. Rev.* **75**, 65–90 (2017).
108. A. M. Liberman, F. S. Cooper, D. P. Shankweiler, M. Studdert-Kennedy, Perception of the speech code. *Psychol. Rev.* **74**, 431–461 (1967).
109. B. Galantucci, C. A. Fowler, M. T. Turvey, The motor theory of speech perception reviewed. *Psychon. Bull. Rev.* **13**, 361–377 (2006).
110. R. Möttönen, K. E. Watkins, Using TMS to study the role of the articulatory motor system in speech perception. *Aphasiology* **26**, 1103–1118 (2012).
111. G. Hickok, L. L. Holt, A. J. Lotto, Response to Wilson: What does motor cortex contribute to speech perception? *Trends Cogn. Sci.* **13**, 330–331 (2009).
112. National Research Council, *Guide for the Care and Use of Laboratory Animals*, (National Academies Press, Washington, DC, 8th Ed., 2011).
113. M. Ortiz-Rios *et al*., Functional MRI of the vocalization-processing network in the macaque brain. *Front. Neurosci.* **9**, 113 (2015).
114. L. L. Jackson, R. S. Heffner, H. E. Heffner, Free-field audiogram of the Japanese macaque (*Macaca fuscata*). *J. Acoust. Soc. Am.* **106**, 3017–3023 (1999).
115. P. Kuśmierek, J. P. Rauschecker, Functional specialization of medial auditory belt cortex in the alert rhesus monkey. *J. Neurophysiol.* **102**, 1606–1622 (2009).
116. D. A. Hall *et al*., "Sparse" temporal sampling in auditory fMRI. *Hum. Brain Mapp.* **7**, 213–223 (1999).