The Neural Basis of Vocal Pitch Imitation in Humans

Michel Belyk[1], Peter Q. Pfordresher[2], Mario Liotti[3], and Steven Brown[1]

Abstract

■ Vocal imitation is a phenotype that is unique to humans among all primate species, and so an understanding of its neural basis is critical in explaining the emergence of both speech and song in human evolution. Two principal neural models of vocal imitation have emerged from a consideration of nonhuman animals. One hypothesis suggests that putative mirror neurons in the inferior frontal gyrus pars opercularis of Broca's area may be important for imitation. An alternative hypothesis derived from the study of songbirds suggests that the corticostriate motor pathway performs sensorimotor processes that are specific to vocal imitation. Using fMRI with a sparse event-related sampling design, we investigated the neural basis of vocal imitation in humans by comparing imitative vocal production of pitch sequences with both nonimitative vocal production and pitch discrimination. The strongest difference between these tasks was found in the putamen bilaterally, providing a striking parallel to the role of the analogous region in songbirds. Other areas preferentially activated during imitation included the orofacial motor cortex, Rolandic operculum, and SMA, which together outline the corticostriate motor loop. No differences were seen in the inferior frontal gyrus. The corticostriate system thus appears to be the central pathway for vocal imitation in humans, as predicted from an analogy with songbirds. ■

INTRODUCTION

Although most vertebrates have the capacity to vocalize, very few species have the ability to learn their species-specific vocal repertoires through vocal imitation, where vocal imitation is defined as the reproduction of previously experienced auditory events (Mercado, Mantell, & Pfordresher, 2014). Among the principal mammalian exceptions are humans, dolphins (King & Sayigh, 2013), whales (Noad, Cato, Bryden, Jenner, & Jenner, 2000), and bats (Knörnschild, Nagy, Metz, Mayer, & von Helversen, 2010). Limited evidence also suggests that elephants (Stoeger et al., 2012; Poole, Tyack, Stoeger-Horwath, & Watwood, 2005), seals (Sanvito, Galimberti, & Miller, 2007; Ralls, Fiorelli, & Gish, 1985), and mice (Arriaga & Jarvis, 2013) may be capable of vocal imitation. This list of species is notably lacking in nonhuman primates. Looking beyond the mammalian class, three lineages of birds—namely, parrots, hummingbirds, and songbirds—are capable of learning to produce novel sounds through vocal imitation (Nottebohm, 1972). Vocal imitation in humans is important not only during childhood development for the establishment of large and flexible acoustic repertoires for speech and music (Trehub, 2001; Studdert-Kennedy, 2000; Kuhl & Meltzoff, 1996; Papousek, 1996; Poulson, Kymissis, Reeve, Andreators, & Reeve, 1991) but also throughout adult life for the ability to, for example, learn musical melodies and produce the sounds of a foreign language.

Although theories of vocal imitation are diverse, they tend to agree on a core set of processes related to the sensorimotor translation of perceived sounds (Pfordresher et al., 2015; Pfordresher & Mantell, 2014; Hutchins & Moreno, 2013; Berkowska & Dalla Bella, 2009; Dalla Bella & Berkowska, 2009). As shown in Figure 1, vocal imitation requires that an individual perceive a target sound, map the acoustic properties of the target onto phonatory and articulatory motor commands through a process of inverse modeling, and then execute those commands to vocally reproduce the target sound. Inverse models involve the use of an internal model of sensorimotor relationships (Kawato, 1999) based on learned associations between motor activity and sensory stimulation (Hanuschkin, Ganguli, & Hahnloser, 2013). Inverse models provide a mechanism for the classic ideomotor principle of motor planning (James, 1890; cf. Shin, Proctor, & Capaldi, 2010) whereby motor plans are configured with reference to their anticipated outcomes. Inverse models are a key component of vocal imitation in that they allow imitators to plan motor movements that are based, for example, on pitch patterns that they have not previously produced (Pfordresher & Mantell, 2014).

There is a widespread population of individuals—colloquially known as "tone deaf" individuals, but more accurately described as "poor-pitch singers"—who have a specific deficit in the sensorimotor translation involved in vocal pitch imitation. Poor-pitch singers are often accurate at encoding auditory stimuli—as demonstrated by

[1]McMaster University, Hamilton, Canada, [2]State University at Buffalo, New York, [3]Simon Fraser University, Burnaby, Canada

doi:10.1162/jocn_a_00914

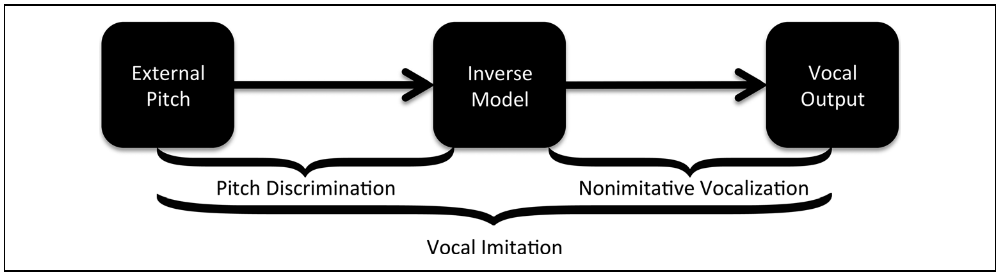

Figure 1. Model of vocal pitch imitation. In vocal imitation, an external pitch stimulus is perceived, converted to a motor code via an inverse model, and this motor program is then executed at the level of the larynx.

performance on pitch discrimination tasks (Pfordresher & Brown, 2007)—but are deficient in translating that internal model into an appropriate motor signal so as to match the acoustic properties of the model (Pfordresher & Mantell, 2014). Their deficit is thus neither sensory nor motor, but instead sensorimotor (i.e., imitative). This is suggestive of a specific deficit in mapping auditory percepts onto phonatory motor commands.

Although there are few neural models of the general capacity for vocal imitation in humans, neural models of speech processing may describe similar processes. Indeed, models of singing are neuroanatomically similar to models of speech production (e.g., Loui, 2015). This follows from the involvement of a common audio–vocal network in speech and nonspeech vocalization (Belyk & Brown, 2015; Grabski et al., 2012, 2013; Chang, Kenney, Loucks, Poletto, & Ludlow, 2009; Brown, Ngan, & Liotti, 2008; Olthoff, Baudewig, Kruse, & Dechent, 2008; Jeffries, Fritz, & Braun, 2003). One early neural model of speech based on neurological observations of aphasic patients dealt explicitly with speech repetition, a form of vocal imitation, in humans. The classic Wernicke–Geschwind model (Geschwind, 1970) posits that auditory information is transmitted from the posterior part of the superior temporal gyrus (pSTG) to the inferior frontal gyrus (IFG) via the arcuate fasciculus (AF) and then presumably to the motor cortex for vocal execution, although the model does not specify this final step. Lesions to the AF, which effectively disconnect the pSTG from the IFG, cause deficits specific to vocal imitation, with spared speech comprehension and otherwise fluent speech production. A similar association has been described between AF integrity and singing (Loui, Alsop, & Schlaug, 2009).

Modern models of speech perception similarly posit an AF-mediated audio–motor linkage (e.g., Hickok, Houde, & Rong, 2011; Rauschecker & Scott, 2009; Hickok & Poeppel, 2007). In particular, Hickok and Poeppel's "dorsal stream" is proposed to mediate "the acquisition of new vocabulary" (p. 399), which is a form of vocal learning. Models of speech production deal more explicitly with the transfer of auditory information to the motor system. For example, Warren, Wise, and Warren (2005) proposed that posterior temporal areas sequence auditory features, whereas Rauschecker (2014) proposed that such auditory sequences are stored in premotor brain areas, allowing them to be later reproduced by the motor system. Guenther and Vladusich (2012) posited that feedback mechanisms contribute to imitation by iteratively modifying speech sound maps in the IFG after repeated attempts at producing a novel sound.

Neural models of vocal imitation have taken their lead from theories of gestural imitation based on mirror neurons. Mirror neurons are cells that have been described in the brains of monkeys that fire both when an animal perceives and produces a particular action (Di Pallegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992). Although the single-cell recording studies necessary to demonstrate the existence of mirror neurons in the human brain have not been conducted, neuroimaging studies have identified brain areas that constitute populations of cells that together display mirror-like properties (Gazzola & Keysers, 2009). Among these putative mirror neuron regions is the posterior portion of Broca's area, consisting of Brodmann's area 44 (BA 44) in the IFG pars opercularis. This region is activated both when viewing manual gestures and when producing them from memory (Iacoboni et al., 1999). However, activation is greatest when imitating novel gestures, suggesting a specific role for this area in gestural imitation. Although a meta-analysis of the gestural imitation literature questioned the reliability of such an imitation effect in the IFG pars opercularis (Molenberghs, Cunnington, & Mattingley, 2009), repetitive TMS of this region disrupts manual imitation (Heiser, Iacoboni, Maeda, Marcus, & Mazziotta, 2003). Such findings have led researchers to speculate that the IFG may also be a key region for vocal learning via imitation (Iacoboni et al., 1999; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996).

Certain species of birds that possess the capacity for vocal production learning provide an alternative neural model for vocal imitation. In contrast to monkeys, three lineages of birds—namely, parrots, hummingbirds, and songbirds—are capable of learning novel vocalizations through vocal imitation (Nottebohm, 1972). The vocal system of vocally imitating birds, particularly songbirds, has been studied extensively (Jarvis et al., 2005). The avian song system consists of two pathways: a descending vocal–motor pathway and a forebrain–striatal loop. Although lesions to the descending vocal–motor pathway profoundly disrupt song production (Nottebohm, Stokes, & Leonard, 1976), lesions to the forebrain–striatal loop disrupt vocal imitation and song learning but spare the production of songs that have already been learned (Sohrabji, Nordeen, & Nordeen, 1990; Bottjer, Miesner, & Arnold, 1984). Neurophysiological evidence suggests that neurons along the forebrain–striatal loop compute inverse models that map target sounds onto the motor commands that reproduce them (Giret, Kornfeld, Ganguli, & Hahnloser, 2014). The brain areas that comprise the two songbird vocal pathways have analogues in the human brain (see Jarvis et al., 2005, for a review), and these analogues are also active when humans sing (Brown, Martinez, Hodges, Fox, & Parsons, 2004). Indeed Area X, a key node in the songbird forebrain–striatal loop, shares molecular specializations with the human putamen (Pfenning et al., 2014). Although the BG as a whole are highly conserved across vertebrates, species may develop novel modules as they evolve new behaviors (Grillner, Robertson, & Stephenson-Jones, 2013). Hence, one hypothesis is that humans and songbirds may have convergently evolved novel modules in the BG that support vocal imitation.

Vocal imitation of pitch is an ideal medium for examining audio–vocal matching because pitch is a highly salient component of vocal communication that can be measured with greater simplicity and precision than either gestural or articulatory imitation. Pitch varies along a single dimension whose relation to the acoustic property of fundamental frequency is well known and therefore lends itself to empirical measurement. Two neuroimaging studies of vocal pitch imitation (Garnier, Lamalle, & Sato, 2013; Brown et al., 2004) and several studies of speech repetition have observed that vocal imitation activates a suite of brain areas that constitute the audio–vocal system, including both the IFG and BG (Segawa, Tourville, Beal, & Guenther, 2015; Mashal, Solodkin, Dick, Elinor Chen, & Small, 2012; Reiterer et al., 2011; Peeva et al., 2010; Rauschecker, Pringle, & Watkins, 2008). However, this network is commonly activated during vocal–motor tasks in general (Grabski et al., 2013; Brown et al., 2009; Simonyan, Ostuni, Ludlow, & Horwitz, 2009; Olthoff et al., 2008; Loucks, Poletto, Simonyan, Reynolds, & Ludlow, 2007; Jeffries et al., 2003).

This study attempted to compare imitative vocalization with the highly matched control conditions of nonimitative vocalization and pitch discrimination using sparse temporal sampling (Hall et al., 1999) so as to measure behavioral performance in the scanner. The principal aim was to shed light on the unique ability of humans among primates to perform vocal imitation by comparing the two competing hypotheses that either the IFG or the corticostriate system supports vocal imitation in humans, as predicted by the "gestural imitation" and "avian song system" animal models, respectively. In the imitation condition, participants listened to novel melodies and then imitated them vocally, thereby engaging all of the processes shown in Figure 1. In a nonimitative vocalization condition, participants were visually cued to sing highly familiar melodies, thereby engaging preexisting motor commands. Finally, in a pitch discrimination condition, participants heard pitch sequences and had to detect pitch changes, thereby engaging auditory but not vocal–motor processes.

METHODS

Participants

Thirteen participants (median age = 24 years, range = 19–48 years, 6 women, 1 left-handed) were recruited at Simon Fraser University. A 14th participant was excluded because of undiagnosed hydrocephalus. Participants were prescreened to verify that they were accurate vocal imitators using stimuli similar to those used in the vocal imitation task in this study (see below). Four additional participants were excluded after prescreening. The 13 remaining participants had absolute note errors of less than one semitone (i.e., 100 cents), on average (see Imitation Analysis section), which was the criterion for accurate imitation established in Pfordresher and Brown (2007). Pfordresher, Brown, Meier, Belyk, and Liotti (2010) estimated that approximately 87% of the population exceeds this criterion. All participants provided written informed consent before prescreening. The experimental protocol was approved by the research ethics board of Simon Fraser University.

Stimuli and Procedure

Participants completed each of three tasks twice in separate runs in random order. For each task, the same stimuli were presented across runs, but in counterbalanced pseudorandom order. Each experimental task consisted of a visual cue, a four-note auditory stimulus, a response period, and a variable delay before image acquisition (Figure 2). Experimental trials alternated with a baseline condition, during which participants fixated on a crosshair. The eyes were kept open in all scans.

The primary task of interest was a vocal pitch imitation task, during which participants listened to and then repeated short melodies. Two control conditions (i.e., nonimitative vocalization and pitch discrimination) sought to match the motor and sensory demands, respectively, of the vocal imitation task.

Vocal Imitation Task

Eighteen novel four-note melodies were synthesized in a vocal timbre on the vowel /u/ using Vocaloid (Leon, Zero-G Limited, Okehampton, UK). All melodies were isochronous with 600-msec interonset intervals, with a 50-msec 10-dB fade-in and drop-off. Notes ranged from A_2 (110 Hz) to E_3 (164.81 Hz) for men and from A_3 (220 Hz) to E_4 (329.63 Hz) for women. Stimuli were generated in equal numbers with three levels of complexity, in accordance with the stimuli of Pfordresher and Brown (2007). "Note" stimuli consisted of a sequence of four identical notes. "Interval" stimuli consisted of two doublets

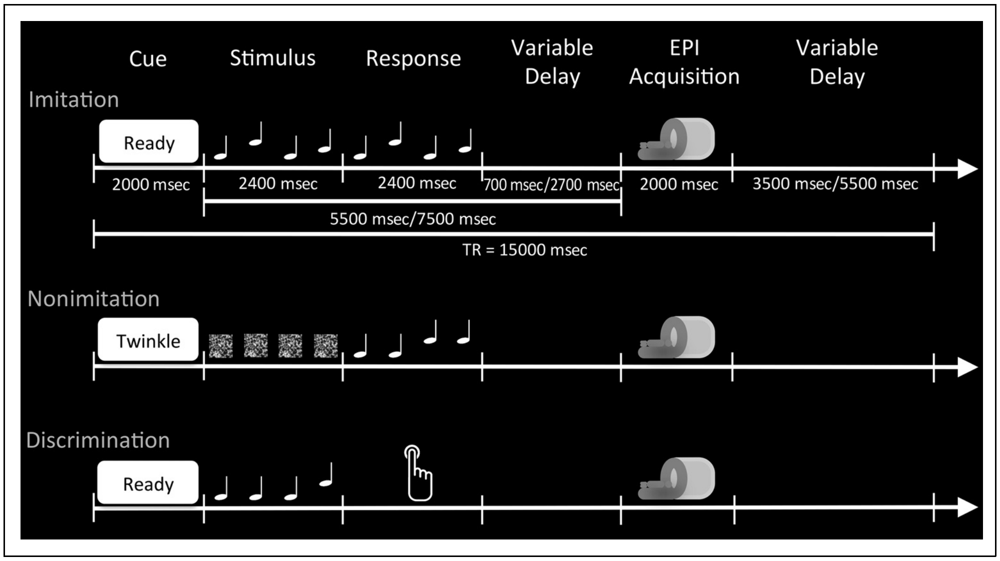

Figure 2. Trial timing. The timing of trials within each of the three conditions is depicted. In the vocal imitation condition, participants heard novel four-note melodies and then imitated them vocally. In the nonimitative vocalization condition, participants were visually cued with the name of a highly familiar melody, heard four task-irrelevant white noise bursts, and then sang the first four notes of the target melody. In the pitch discrimination condition, participants heard a series of three identical notes followed by a fourth note, and then indicated on a response pad whether the fourth note was the same or different than the preceding three. On the basis of the use of a sparse temporal sampling design, EPI images were collected after each trial. Hence, participants performed all tasks in the absence of scanner noise.

of notes with a single interval between the first and second doublet (e.g., "AAEE"). "Melody" stimuli consisted of a series of nonrepeating notes (e.g., ABC#E). A "Ready" screen was displayed 2 sec before the onset of a stimulus to indicate that a trial was about to begin. The target melody was presented for 2400 msec followed by a 2400-msec response period, during which participants were instructed to imitate the target melody.

Nonimitative Vocalization Task

Participants were visually cued with the name of a familiar melody and instructed to vocalize the first four notes of the melody. Participants vocalized either a monotone sequence (i.e., four identical pitches), "Twinkle, Twinkle," or "Mary Had a Little Lamb." These stimuli matched the number of note changes in the note, interval, and melody stimuli, respectively, of the vocal imitation task. After the verbal cue, four white noise bursts were presented that matched the amplitude and duration of the stimulus melodies of the vocal imitation task. This was done to match the level of auditory stimulation that was present in the vocal imitation and pitch discrimination conditions. Participants were instructed to produce the familiar melodies from memory in a comfortable part of their vocal range after the white noise bursts were completed.

Pitch Discrimination Task

Eighteen four-note melodies were synthesized in the same manner as the target melodies of the vocal imitation task. The first three notes of each melody were A_2 for men or A_3 for women. On half of the trials, the final note was identical to the initial notes. In the remaining trials, the final note was 25, 50, 100, 200, 400, or 600 cents higher or lower than the initial notes (where 100 cents = 1 equal-tempered semitone). Participants pressed a button to indicate whether the final note was identical or not to the initial notes. Button presses were recorded on an MRI-compatible button box with the index and middle fingers of the right hand, where the "same" and "different" options were counterbalanced across participants.

Imitation Analysis

Sung melodies were recorded from participants in the scanner using an MRI-compatible microphone that fed into the Avotek patient communication system, itself connected to a laptop computer running Adobe Audition (San Jose, CA). Sung melodies from the scanner were then subjected to acoustic analysis. The pitch of each sung note was extracted using the autocorrelation algorithm in Praat (Boersma & Weenink, 2011) and compared with the corresponding notes of each target melody. The intervals of the target and sung melodies were calculated as the difference between adjacent notes in the target and sung melodies, respectively. Performance on the vocal imitation task in the scanner was assessed by both the accuracy and precision of both the notes and the melodic intervals, as described in Pfordresher et al. (2010). Accuracy was measured as the mean signed difference between the notes or intervals of sung melodies and those of the target melodies averaged across pitch classes. Precision was measured as the standard deviation of note and interval errors across pitch classes.

MRI

MRIs were acquired with a Phillips 3-T MRI. Functional images sensitive the BOLD signal were collected with gradient-echo sequences according to a sparse event-related sampling design (Hall et al., 1999). Samples were collected 5500 or 7500 msec after stimulus onset on alternating trials to eliminate scanner noise during auditory stimulus presentation and vocalization as well as to minimize movement-related artifacts during image acquisition. These jittered acquisition times were selected to ensure that data were collected around the expected maxima of the BOLD response after accounting for the hemodynamic lag. Imaging parameters were as follows: repetition time = 15000 msec, acquisition time = 2000 msec, echo time = 33 msec, flip angle = 90°, 36 slices, slice thickness = 3 mm, gap = 1 mm, in-plane resolution = 1.875 mm × 1.875 mm, matrix = 128 × 128, and field of view = 240 mm. A total of 39 whole-brain volumes were collected per scan. The first three were discarded, leaving 36 volumes, corresponding to 18 alternations between task and rest trials. A T1-weighted image with 1-mm isotropic voxels and field of view of 256 mm × 256 mm × 170 mm was also collected for image registration.

Image Analysis

Each functional scan was spatially smoothed with a Gaussian kernel of 4 mm FWHM. High-pass filtering was accomplished by modeling the low-frequency components of the sparse time series with a general linear model with a basis set of one linear, two sine, and two cosine functions. The estimates of this model were subtracted from the sparse time series to remove the influence scanner drift. Each sample was spatially realigned with the first sample in its run to correct for head motion. Translational and rotational corrections did not exceed an acceptable level of 1.5 mm and 1.5°, respectively, for any participant. Following realignment, each functional scan was normalized to the Talairach template (Talairach & Tournoux, 1988).

In a first-level fixed-effects analysis, beta weights associated with a simple task versus rest model were fitted to the observed BOLD signal time course in each voxel for each participant using the general linear model, as implemented in (Brain Voyager QX 2.8, Maastricht, The Netherlands). Six head motion parameters describing translation and rotation of the head were included as nuisance regressors. Because image acquisition began either 5500 msec or 7500 msec after task onset as part of the sparse-imaging design, no hemodynamic transformation was applied to the statistical model. The raw BOLD signal was transformed to percent signal change for group analyses. Contrast images for each task-versus-rest contrast were brought forward into a second-level random effects analysis. Talairach coordinates were extracted for all contrasts using NeuroElf (neuroelf.net), and activations were labeled according to the atlas of Talairach and Tournoux (1988) and verified against the anatomy of individual participants.

Statistical Contrasts

To localize the basic audio–vocal network, we performed a three-way conjunction between vocal imitation, nonimitative vocalization, and pitch discrimination. To further identify vocal-motor-related activations, we performed a conjunction of the contrasts [Imitation > Discrimination] ∩ [Nonimitation > Discrimination]. Because a strong BOLD response was expected from these motor and auditory tasks relative to rest, these contrasts were corrected for multiple comparisons with an overly conservative false discovery rate (FDR) of $q < 0.01$ and an additional arbitrarily selected cluster threshold of $k >$ 12 voxels to eliminate small clusters.

To identify regions of the vocal network that were preferentially activated by vocal imitation, we performed a conjunction of the contrasts [Imitation > Nonimitation] ∩ [Imitation > Discrimination]. This conjunction identified brain regions that were more active during vocal imitation than both the nonimitative vocalization and pitch discrimination control conditions. Because these high-level contrasts compared highly matched conditions, a more sensitive threshold was applied that still corrected for multiple comparisons. A cluster-wise error rate of $p <$ .05 was maintained by combining an uncorrected voxel-wise threshold of $p < .05$ with a cluster threshold of $k >$ 18 voxels, as determined by Monte Carlo simulation.

ROI Analysis

We identified functionally localized ROIs averaged across the volume of 5-mm cubes drawn around the activation peaks of each brain region identified in the vocal imitation conjunction analysis. Beta coefficients from first-level analyses were extracted for each participant from each brain area for each condition. An examination of these data revealed that the left-handed participant was not a statistical outlier.

RESULTS

Behavioral Data

The mean accuracy score of vocal imitation performance in the scanner, combined across note and interval measurements, was 44.5 cents (*SD* = 17.0). The mean precision of imitation was 66.4 cents (*SD* = 41.6). This suggests that the participants were accurate and precise imitators, according to established criteria for these parameters (Pfordresher et al., 2010; Pfordresher & Brown, 2007). These measurements replicated imitation performance during the prescreening experiments. Median performance on the pitch discrimination task was 94.4%.

Imaging Data

Vocal imitation, nonimitative vocalization, and pitch discrimination all activated a basic audio–vocal network. A conjunction between these three conditions (Figure 3) revealed shared activations in bilateral Heschl's gyrus (BA 41) extending into the pSTG (BA 42 and 22), orofacial premotor cortex (BA 6), IFG pars opercularis (BA 44), anterior insula (BA 13), putamen, thalamus, and lateral cerebellum. Shared activations were observed in the bilateral SMA, ACC, and cerebellar vermis. The shared audio–vocal areas identified in this conjunction reflect a neural system for the internal encoding of melodies resulting from either online perception or access from long-term stores (Table 1).

A conjunction of the contrasts [Imitation > Discrimination] ∩ [Nonimitation > Discrimination] revealed a set of regions preferentially activated during vocal production. This extended the abovementioned network to include the bilateral orofacial motor cortex and Rolandic operculum as well as bilateral Heschl's gyrus (BA 41) and right SMA (Table 2).

Vocal Imitation

The conjunction [Imitation > Nonimitation] ∩ [Imitation > Discrimination] revealed a subset of the audio–vocal network that was more active during vocal imitation than both nonimitative vocalization and pitch discrimination (Figure 4). These areas included the right orofacial sensorimotor cortex (BA 4/3), left subcentral gyrus (BA 6/43), the bilateral SMA (BA 6), and bilateral putamen. Notably, the IFG was not among the areas revealed by this analysis. All of these areas were also present in each condition individually (as seen in Figure 3), suggesting that, although they were preferentially engaged by vocal imitation, they were by no means specific to that task. Descriptive ROI plots (Figure 5) of these regions indicated that they were activated in all three tasks—not just the imitation task—suggestive of a potential species difference from the songbird.

Partial correlations between the level of activation, as indexed by beta coefficients in first-level analyses, and mean absolute error in the vocal imitation task, controlling for age at the time of the scan, did not reach significance for any ROI. The coefficients of partial regression were $r(11) = 0.18, p = .56$ for the left striatum, $r(11) = 0.12$, $p = .70$ for the right striatum, $r(11) = -0.21$, $p = .50$ for the right Rolandic operculum, $r(11) = -0.47, p = .09$ for the left M1, $r(11) = -0.25, p = .41$ for the SMA (Table 3).

DISCUSSION

To shed light on the unique ability of humans among primates to perform vocal imitation, we conducted a targeted comparison between imitative vocalization and the closely matched tasks of nonimitative vocalization and auditory discrimination so as to identify brain areas preferentially activated by imitation. We did so using accurate imitators and a sparse temporal sampling fMRI protocol that both created a silent environment for the participants to perform the task and that permitted us to record vocal behavior in the scanner. The results failed to show a significant imitative effect in the IFG but instead demonstrated a clear, though small, effect in the corticostriate pathway, including the putamen, SMA, and orofacial sensorimotor cortex, suggesting that these regions are preferentially engaged during vocal imitation. Although the degree of activation in these areas did not correlate with imitative performance in the scanner, this

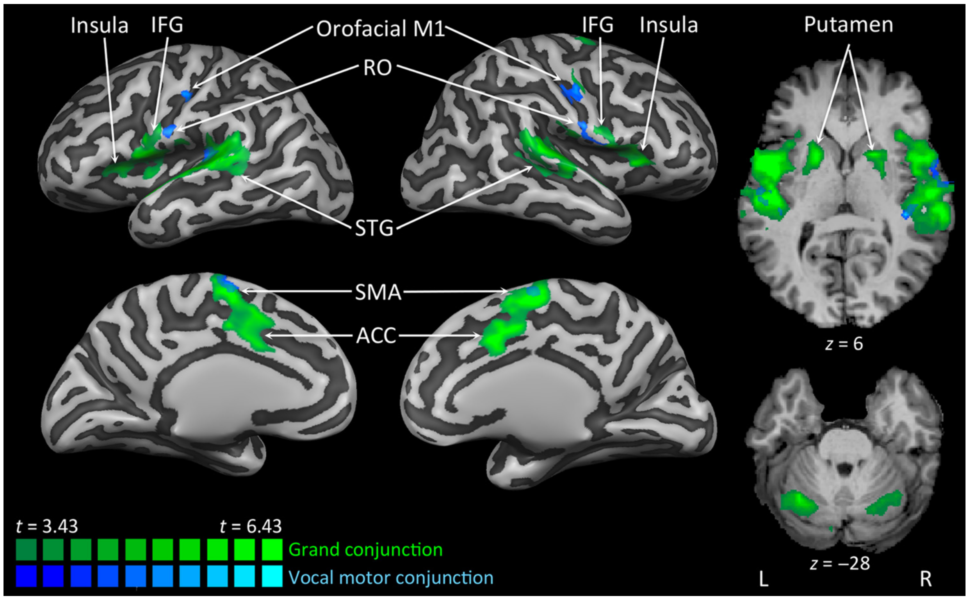

Figure 3. The audio–vocal network. Activation maps for the conjunction of vocal imitation, nonimitative vocalization, and pitch discrimination (green) show those elements of the audio–vocal system that are activated during all three tasks. The conjunction of contrasts [Imitation > Discrimination] ∩ [Nonimitation > Discrimination] (blue) shows brain areas that were preferentially engaged during vocalization. Both maps were thresholded at FDR $q < 0.01$ $k > 12$. M1 = primary motor cortex; RO = Rolandic operculum.

Table 1. Low-level Contrasts

	Imitation						*Nonimitation*						*Discrimination*					
Brain Regions	*x*	*y*	*z*	*Voxels*	*mm^3*	*t*	*x*	*y*	*z*	*Voxels*	*mm^3*	*t*	*x*	*y*	*z*	*Voxels*	*mm^3*	*t*
Frontal Lobe																		
SMA (BA 6)	3	−7	59	422	5934	36.0	3	−7	59	261	3670	19.9	7	−7	56	35	492	11.6
ACC (BA 32)							3	11	37	71	998	12.8	3	10	42	483	6792	13.6
Pericentral (BA 6/4/3)	−54	−6	23	45	633	17.7	−50	−14	23	30	422	11.8	−46	−5	18	40	563	14.0
	−44	−18	41	51	717	9.8	−44	−15	46	50	703	7.8	−49	−30	37	45	633	10.6
	53	−19	42	98	1378	10.6	57	−11	40	45	633	9.8	−53	−22	17	349	4908	13.2
	3	−45	66	32	450	7.0							−58	−25	29	56	788	12.8
Anterior insula (BA 13)	−37	17	19	25	352	9.5							−40	17	16	44	619	10.1
IFG (BA 44)	−52	1	12	190	2672	23.6	−49	1	12	195	2742	21.3	−52	4	12	239	3361	20.4
	57	−1	10	523	7355	24.6							36	16	14	260	3656	15.3
MFG													44	34	33	27	380	8.3
Temporal Lobe																		
Heschl's gyrus (BA 41)	−48	−22	13	632	8888	29.5	45	−24	8	27	380	19.5	−50	−33	17	73	1027	11.7
STG (BA 42)	−42	−39	19	38	534	25.7	−58	−25	15	485	6820	29.8						
	59	−28	12	236	3319	23.5	59	−19	10	558	7847	26.0	59	−28	19	263	3698	13.1
STG (BA 22)							51	−10	9	35	492	17.9						
Parietal Lobe																		
IPL (BA 40)							0	−51	65	48	675	8.3	−38	−48	54	50	703	10.2
													−41	−49	47	119	1673	10.0
													48	−47	33	115	1617	10.0
PCC (BA 23)													−6	−23	26	101	1420	9.8
Subcortical																		
BG	−23	10	16	214	3009	15.3	−19	7	13	28	394	8.0	−23	3	13	225	3164	12.4
							−19	−2	22	29	408	7.6	31	−5	9	41	577	11.2
							−16	−5	6	78	1097	9.9	22	−1	15	83	1167	10.6
	16	−1	2	195	2742	13.0	16	−4	6	40	563	10.0						
Thalamus	15	−25	1	40	563	11.3	15	−25	1	40	563	11.3	−19	−16	15	25	352	7.4
Cerebellum	0	−75	−20	52	731	11.8	0	−78	−17	110	1547	12.6	−3	−53	−5	51	717	8.7
	−3	−63	−5	53	745	11.4	−28	−57	−17	72	1013	10.2	−39	−48	−23	66	928	8.6
	−27	−54	−17	29	408	8.5	21	−59	−19	95	1336	10.7	33	−51	−21	40	563	7.3
	36	−48	−23	53	745	9.7	44	−59	−24	132	1856	10.8						
	12	−59	−13	54	759	9.6												

Location of peak voxels for the three experimental contrasts against fixation. After each anatomical name in the brain region column, the Brodmann's number for that region is listed. The columns labeled as *x*, *y*, and *z* contain the Talairach coordinates for the peak of each cluster reaching significance at '$q < 0.01$ with cluster threshold $k > 12$. IPL = inferior parietal lobule; MFG = middle frontal gyrus; PCC = posterior cingulate cortex.

Table 2. The Audio–Vocal Network

	Grand Conjunction						*Vocal–Motor Conjunction*					
Brain Regions	*x*	*y*	*z*	*Voxels*	*mm^3*	*t*	*x*	*y*	*z*	*Voxels*	*mm^3*	*t*
Front Lobe												
Orofacial motor cortex (BA 4)							−39	−19	40	23	323	4.3
							51	−10	46	42	591	4.3
Rolandic operculum (6/43)							−57	−10	22	21	295	4.2
							57	−7	16	26	366	4.2
Precentral gyrus (BA 6)	48	−4	49	37	520	4.3						
Anterior insula (BA 13)	−30	20	16	43	605	4.6						
	36	23	10	211	2967	4.4						
IFG pars opercularis (BA 44)	−51	5	7	431	6061	5.0						
	54	8	4	83	1167	5.4						
IFG pars opercularis (BA 44/6)	57	2	19	13	183	4.7						
SMA (BA 6)	6	−7	61	460	6469	4.8	−6	−7	67	35	492	4.2
ACC (BA 32)	3	11	40	44	49	619	6.5					
Temporal Lobe												
Heschl's gyrus (BA 41)							−48	−19	10	20	281	4.2
							−39	−28	13	43	605	4.5
	54	−16	10	21	295	6.4	39	−28	7	17	239	4.1
pSTG (BA 42)	−54	−31	16	503	7073	5.2						
	63	−28	10	495	6961	4.7						
Parietal Lobe												
Postcentral gyrus (BA 40)	−57	−19	16	25	352	8.5						
Subcortical												
Striatum	−18	8	10	159	2236	4.5						
	18	11	7	95	1336	4.6						
	−18	2	−5	29	408	4.7						
	15	2	−5	24	338	4.5						
Thalamus	−12	−7	13	21	295	3.8						
Cerebellar hemisphere	−30	−55	−26	89	1252	4.4						
	33	−49	−32	125	1758	4.0						
Cerebellar vermis	−3	−61	−11	80	1125	4.1						

Location of peak voxels for the grand conjunction of vocal imitation, nonimitative vocalization, and pitch discrimination showing those elements of the audio–vocal system that are activated during all three tasks. The conjunction of contrasts [Imitation > Discrimination] ∩ [Nonimitation > Discrimination] shows brain areas that were preferentially engaged during vocalization. Both conjunctions were thresholded at FDR $q < 0.01$ $k > 12$.

may be due to the narrow range of vocal imitation scores in this group of participants, because they were selected on the basis of accurate imitative performance during prescreening.

These results are consistent with an extensive literature showing that the BG function in the acquisition of novel motor sequences (Shmuelof & Krakauer, 2011). Importantly, ROI analyses showed that the putamen was activated both when perceiving pitches and when singing them, hence creating an important link between these two phases of vocal imitation. Consistent with previous research (Garnier et al., 2013; Grabski et al., 2013; Olthoff et al., 2008; Brown & Martinez, 2007; Reiterer, Erb, Grodd, & Wildgruber, 2007; Reiterer et al., 2005), all three tasks, including the nonvocal pitch discrimination task, activated an overlapping set of brain regions that contained the majority of areas comprising the audio–vocal network. Only the orofacial motor cortex and Rolandic operculum, adjacent to the subcentral gyrus phonation area described by Bouchard, Mesgarani, Johnson, and Chang (2013), were specifically activated during vocal production.

The classical model of vocal imitation in humans, namely, the Wernicke–Geschwind model (Geschwind, 1970), implicates the IFG as a key node in the imitative pathway. According to this model, the AF relays auditory information from the temporal lobe to speech-planning areas in the frontal lobe. Lesions to the AF can cause conduction aphasia, characterized by imitation-specific speech deficits, with sparing of both the production and comprehension of speech. The role of the AF in audio–motor integration is also ubiquitous in contemporary models of speech processing. However, we observed no specificity for vocal imitation in the brain areas that lie at either end of the AF (i.e., pSTG and IFG). These findings suggest that, although the AF pathway may be necessary for relaying auditory information to the motor system, processes that are specific to vocal imitation occur downstream of this pathway.

We suggest that one such process is the computation of inverse models in the BG. Stronger activations for imitation compared with nonimitative production were found in several regions of the vocal motor network. Most notably, the putamen, which is analogous to songbird Area X—itself a key node in the vocal imitation pathway of songbirds—was more active during vocal imitation than either nonimitative vocalization or pitch discrimination, although both of these latter tasks also activated the putamen to some degree. This imitation effect is consistent with neurophysiological work in the songbird showing that Area X receives afferents from pallial mirror neurons (Prather, Peters, Nowicki, & Mooney, 2008) and is a strong candidate for being the source of the inverse models that relate target sounds to motor commands (Giret et al., 2014).

To our knowledge only one previous brain imaging study compared vocal pitch imitation with nonimitative vocalizations (Garnier et al., 2013). However, that study failed to detect a main effect of imitation anywhere in the brain, although a correlation with imitative performance was observed in auditory cortex. Speech repetition has been studied more widely and may engage processes similar to vocal pitch imitation. One study

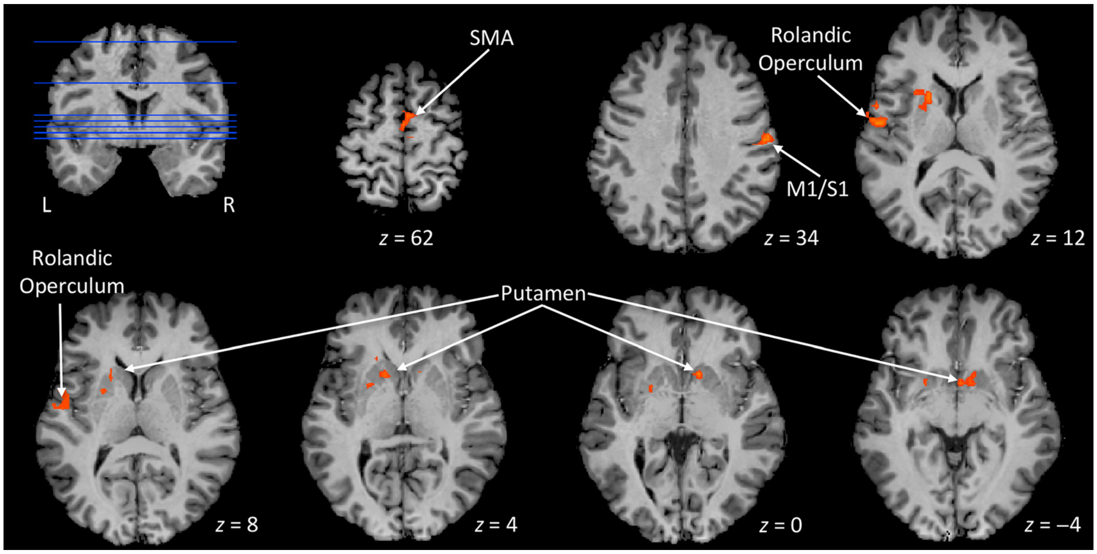

Figure 4. Vocal imitation. A whole-brain map display of the conjunction of high-level contrasts [Imitative vocalization > Nonimitative vocalization] ∩ [Imitation > Discrimination] depicting areas of the brain activated during vocal imitation. A cluster-wise error rate of $p < .05$ was maintained by combining an uncorrected voxel-wise threshold of $p < .05$ with a cluster threshold of $k > 18$ voxels, as determined by Monte Carlo simulation. Blue lines on the coronal slice ($y = 0$) indicate the levels at which axial slices were taken. M1 = primary motor cortex; S1 = primary somatosensory cortex.

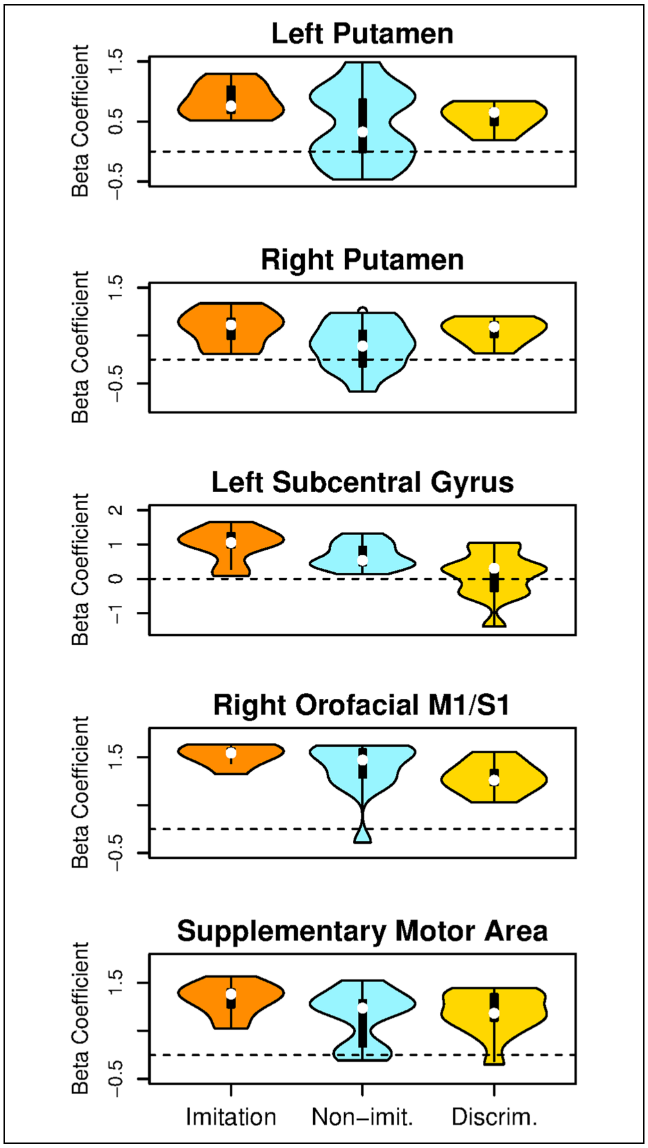

Figure 5. Descriptive ROI plots. Violin plots show the distribution of beta coefficients for imitative vocalization, nonimitative vocalization, and pitch discrimination in each brain area that was preferentially engaged by vocal imitation. The dashed horizontal line marks beta values of zero in each plot. These plots demonstrate that, although vocal imitation preferentially engaged these regions, they were not specific to imitation. This suggests that this corticostriate system contributes to both the encoding and production phases of vocal imitation, in addition to any imitation-specific processes.

observed a correlation between activity in the IFG pars opercularis and ability to repeat foreign words (Reiterer et al., 2011), although another observed increased effective connectivity between the STG and premotor cortex, rather than the IFG, during speech imitation (Mashal et al., 2012). Other studies have observed increased activation of the striatum, including both the putamen and caudate nucleus, when imitating foreign speech sounds (Simmonds, Leech, Iverson, & Wise, 2014). The putamen is activated by pseudoword repetition (Peeva et al., 2010), and the level of activation in the putamen decreases with practice (Rauschecker et al., 2008), consistent with a transition from motor learning to motor program retrieval and the possible continued involvement of the BG in state feedback control (Houde & Nagarajan, 2011). Separate subdivisions of the putamen may underlie imitating novel pseudowords compared with retrieving motor commands to produce well-known real words (Hope et al., 2014). Other areas within the corticostriate loop, including the globus pallidus and pre-SMA, are more active when repeating novel consonant clusters (Segawa et al., 2015).

Figure 6 attempts to summarize the results of this study in the context of the standard model of vocal imitation in the human neuroscience literature, namely, the Wernicke–Geschwind model, which emphasizes the transmission of auditory information from the STG to the IFG via the AF. We argue that this pathway is necessary but not sufficient for vocal imitation to occur. Instead, processing in the putamen beyond that required for either pitch encoding or pitch production alone is needed to match target sounds to vocal motor commands. At present, it is uncertain if the critical connectivity between the BG and the vocal–motor system occurs with the IFG, larynx motor cortex (via the SMA), or both. Further studies of both functional and structural connectivity will be needed to resolve this issue.

Evolutionary Considerations

Comparative neuroscience has revealed evolutionary expansions of brain regions throughout the human audio–vocal system relative to other primates, which has generated several neuroanatomical hypotheses for the evolution of vocal imitation. However, the evolution of vocal imitation is phylogenetically coupled with flexible motor control over the vocal organ, be it a larynx or a syrinx. We are not aware of any species that has the capacity to flexibly produce novel vocalizations in the absence of vocal imitation, or vice versa. Hence, although undoubtedly useful, anatomical comparisons between species necessarily confound adaptations that underlie the sensorimotor transformations required for vocal imitation with sensory or motor adaptations that underlie the capacity for flexible control of the vocal organ. What does seem clear, however, is that the human audio–vocal system evolved the capacity to perform vocal imitation from phylogenetic precursors that lacked both of these abilities.

Several neuro-phenotypical differences have been described between humans and other primates that may be relevant for the emergence of vocal imitation, flexible vocal control, or both. In humans, the AF is more strongly developed than in nonhuman apes (Rilling, Glasser, Jbabdi, Andersson, & Preuss, 2012; Rilling et al., 2008). The IFG pars opercularis contains the evolutionarily novel diagonal

Table 3. Vocal Imitation

	Conjunction of Contrasts					
Brain Regions	*x*	*y*	*z*	*Voxels*	*Size (mm^3)*	*t*
Frontal Lobe						
Orofacial M1/S1 (BA 4/3)	53	−14	36	291	4092	3.78
Rolandic operculum (BA 6/43)	−49	1	6	231	3248	4.27
SMA (BA 6)	−1	−8	63	268	3769	3.18
Subcortical						
Putamen	11	13	0	140	1969	3.13
Putamen	−22	−5	18	358	5034	40.8

Location of peak voxels for the conjunction of high-level contrasts [Imitative vocalization > Nonimitative vocalization] ∩ [Imitation > Discrimination]. After each anatomical name in the brain region column, the Brodmann's number for that region is listed. The columns labeled as x, y, and z contain the Talairach coordinates for the peak of each cluster reaching significance. A cluster-wise error rate of $p < .05$ was maintained by combining an uncorrected voxel-wise threshold of $p < .05$ with a cluster threshold of $k > 18$ voxels, as determined by Monte Carlo simulation. M1 = primary motor cortex; S1 = primary somatosensory cortex.

sulcus, which is associated with increased cortical volume of this area (Keller, Roberts, & Hopkins, 2009). In humans, corticobulbar neurons from the motor cortex project directly to the nucleus ambiguus (Iwatsubo, Kuzuhara, & Kanemitsu, 1990; Kuypers, 1958b), whereas such direct connections are sparse in chimpanzees (Kuypers, 1958a) and absent in monkeys (Jürgens & Ehrenreich, 2007). In addition, the cortical larynx area has undergone an evolutionary migration from the premotor cortex in monkeys (Hast, Fischer, & Wetzel, 1974) to an intermediate position in great apes (Leyton & Sherrington, 1917) to the primary motor cortex in humans (Pfenning et al., 2014; Bouchard et al., 2013; Brown et al., 2008; Loucks et al., 2007). Although comparative neuroscience has greatly advanced our knowledge of brain evolution, such neuroanatomical differences cannot be specifically attributed to the emergence of vocal imitation in humans without further functional evidence.

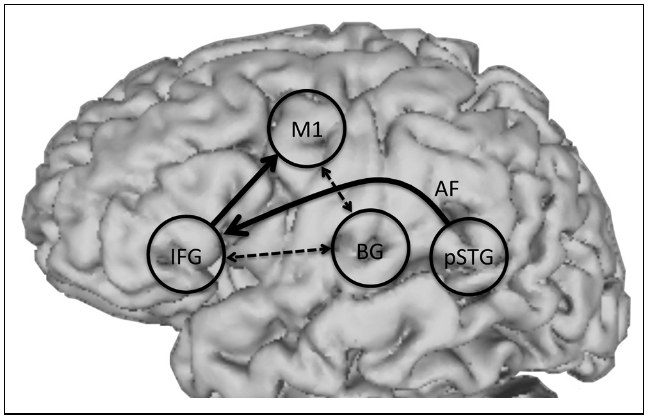

Figure 6. A simple neural model of vocal imitation. The model summarizes the results of this study in the context of pathways common to neural models of speech processing. Target sounds are processed in auditory regions, including the posterior part of the STG, and are transmitted to the frontal lobe along the AF to the IFG, which in turn projects to the primary motor cortex, which executes motor commands to reproduce the target sound. Results from the current study suggest that processing through the corticostriate loop is necessary for matching auditory targets with motor commands. However, it is unclear both from the present experiment and from songbird models of this system whether the anatomical connections of the BG that serve vocal imitation occur at the level of the IFG or motor cortex. This uncertainty is indicated by the dashed lines connecting these structures to the BG.

Some of the critical evidence that comes to bear on the evolution of the vocal system comes not from a consideration of homology with primates but of analogy with other vocal learning species, most notably songbirds. A large body of evidence links songbird Area X—which is a vocally specialized region of the striatum—to imitation (Jarvis, 2007). Furthermore, there are marked anatomical and molecular similarities between the human and songbird vocal systems, which may reflect a process of convergent evolution (Pfenning et al., 2014; Petkov & Jarvis, 2012; Jarvis, 2007). Lesions to Area X and related structures disrupt vocal learning but have little effect on prelearned song (Sohrabji et al., 1990; Bottjer et al., 1984). These structures contain neurons that may compute inverse models that relate target sounds to motor commands (Giret et al., 2014). Inverse models are maximally efficient for motor learning if they generate variable motor commands (Hanuschkin et al., 2013), because variability is required for motor exploration and thus for improvement on subsequent imitative attempts. Ablating output from the forebrain–striatal loop, such that only the posterior descending pathway drives vocalization, results in highly stereotyped song. In contrast, ablating part of the descending pathway, such that only the forebrain–striatal loop drives vocalization, results in a reversion to the oscine equivalent of babbling, which is characterized by a highly variable song (Aronov, Andalman, & Fee, 2008). Song is typically more variable during undirected singing than when it is directed from a male to a female. Increased variability in neural firing along the forebrain–striatal loop during undirected singing (Hessler & Doupe, 1999) results in increased song variability (Kao, Doupe, & Brainard, 2005; Liu & Nottebohm, 2005), and lesioning this pathway prevents such context-dependent changes in song variability to occur (Kao & Brainard, 2006). The forebrain–striatal loop is therefore believed to participate in both generating inverse models to produce new motor programs and in modulating motor variability to facilitate motor exploration and learning.

One gene that links the vocal systems of humans and songbirds is *FOXP2*. Experimental knockdown of *FOXP2* in the juvenile songbird's Area X selectively disrupts vocal imitation (Haesler et al., 2007), and *FOXP2* expression in this region continues to modulate song variability into adulthood (Teramitsu & White, 2006). In humans, *FOXP2* mutations are associated with extensive speech and language deficits (Lai, Fisher, Hurst, Vargha-Khadem, & Monaco, 2001; Hurst, Baraitser, Auger, Graham, & Norell, 1990), including the inability to imitate novel speech sounds, such as pseudowords (Shriberg et al., 2006; Watkins, Dronkers, & Vargha-Khadem, 2002). Patients with *FOXP2* mutations have reduced activation throughout the vocal system, including the putamen, during pseudoword repetition tasks (Liégeois, Morgan, Connelly, & Vargha-Khadem, 2011). The existing literature is broadly consistent with an analogous role of *FOXP2* in humans and songbirds. However, such a conclusion is limited by the necessary reliance on natural experiments in humans. Experimental evidence from the current study further supports the functional analogy between the songbird forebrain–striatal loop and the human corticostriate loop by demonstrating for the first time that the human putamen is preferentially activated during vocal pitch imitation compared with a well-matched nonimitative vocalization task.

The current study demonstrated that, in humans, the putamen is preferentially engaged by vocal imitation, but it is by no means exclusive to imitative processes. This might suggest a potential species difference between humans and songbirds. Indeed, lesions of Area X in songbirds are not believed to impair the production of songs that have already been learned (although see Kubikova et al., 2014; Kao & Brainard, 2006; Hessler & Doupe, 1999), whereas disruption of the BG system in humans leads to strong vocal production deficits. Degenerative diseases of the BG, such as Parkinson's disease, can cause severe forms of dysphonia and articulatory disturbances (Blumin, Pcolinsky, & Atkins, 2004; Canter, 1963). This suggests that, as with BG control of other effectors, the vocal portion of the putamen supports vocal production. The putamen also coactivates with the rest of the vocal system both when vocalizing (Brown et al., 2009) and when discriminating pitch patterns (Brown & Martinez, 2007). This suggests that the BG may have an underappreciated role in nonmotor functions (Kotz, Schwartze, & Schmidt-Kassow, 2009).

The position of the putamen within the human vocal system remains unclear. In songbirds, Area X receives input from a region whose hypothesized human analogue is the IFG (Petkov & Jarvis, 2012). However, evidence for this analogy remains sparse (Pfenning et al., 2014). Alternatively, the human vocal striatum may receive projections from the SMA, which is the dominant source of afferent fibers for corticostriate motor loops supporting other effectors (Alexander, DeLong, & Strick, 1986; Kunzle, 1975). Indeed, in this study, the SMA, and not the IFG, was preferentially engaged by vocal imitation, suggesting that the SMA may be linked with the putamen during vocal imitation. However, diffusion tensor imaging of the human brain suggests that both the IFG (Ford et al., 2013) and SMA project to the putamen (Leh, Ptito, Chakravarty, & Strafella, 2007; Lehéricy et al., 2004). Further research is required to elucidate the anatomical and functional connectivity of the putamen within the vocal motor system.

Conclusions

We report the results of a highly controlled brain imaging study of vocal pitch imitation in humans. Although the tasks of imitating a novel melody and singing a familiar melody from memory robustly activated a common network of vocal areas, imitation was associated with greater activation in a subset of this network, most prominently the putamen. This region is the putative analogue of a critical node in the forebrain–striatal loop for vocal learning in songbirds. These data provide the first evidence that the putamen—but not the IFG—is preferentially engaged during imitative singing in humans, as predicted by functional analogy with songbird Area X.

Acknowledgments

This work was funded by a grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada to S. B. (grant no. 371336) and the National Science Foundation of the United States (BCS-1256964) to P. Q. P.

Reprint requests should be sent to Michel Belyk, Department of Psychology, Neuroscience & Behaviour, McMaster University, 1280 Main St. West, Hamilton, ON, L8S 4M9, Canada, or via e-mail: belykm@mcmaster.ca.

REFERENCES

Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience, 9,* 357–381.

Aronov, D., Andalman, A. S., & Fee, M. S. (2008). A specialized forebrain circuit for vocal babbling in the juvenile songbird. *Science, 320,* 630–634.

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

Belyk, M., & Brown, S. (2015). Pitch underlies activation of the vocal system during affective vocalization. *Social Cognitive and Affective Neuroscience,* 1–11.

Berkowska, M., & Dalla Bella, S. (2009). Acquired and congenital disorders of sung performance: A review. *Advances in Cognitive Psychology, 5,* 69–83.

Blumin, J. H., Pcolinsky, D. E., & Atkins, J. P. (2004). Laryngeal findings in advanced Parkinson's disease. *Annals of Otology, Rhinology and Laryngology, 113,* 253–258.

Boersma, P., & Weenink, D. (2011). *Praat: Doing phonetics by computer [Computer program]*. Version 5.1.29, retrieved 11 March 2010 from http://www.praat.org/.

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

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

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

Brown, S., & Martinez, M. J. (2007). Activation of premotor vocal areas during musical discrimination. *Brain and Cognition, 63,* 59–69.

Brown, S., Martinez, M. J., Hodges, D. A., Fox, P. T., & Parsons, L. M. (2004). The song system of the human brain. *Cognitive Brain Research, 20,* 363–375.

Brown, S., Ngan, E., & Liotti, M. (2008). A larynx area in the human motor cortex. *Cerebral Cortex, 18,* 837–845.

Canter, G. J. (1963). Speech characteristics of patients with Parkinson's disease: I. Intensity, pitch and duration. *Journal of Speech and Hearing Disorders, 28,* 221–229.

Chang, S. E., Kenney, M. K., Loucks, T. M. J., Poletto, C. J., & Ludlow, C. L. (2009). Common neural substrates support speech and nonspeech vocal tract gestures. *Neuroimage, 47,* 314–325.

Dalla Bella, S., & Berkowska, M. (2009). Singing proficiency in the majority: Normality and "phenotypes" of poor singing. *Annals of the New York Academy of Sciences, 1169,* 99–107.

Di Pallegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: A neurophysiological study. *Experimental Brain Research, 91,* 176–180.

Ford, A. A., Triplett, W., Sudhyadhom, A., Gullett, J., McGregor, K., Fitzgerald, D. B., et al. (2013). Broca's area and its striatal and thalamic connections: A diffusion-MRI tractography study. *Frontiers in Neuroanatomy, 7,* 1–8.

Garnier, M., Lamalle, L., & Sato, M. (2013). Neural correlates of phonetic convergence and speech imitation. *Frontiers in Psychology, 4,* 1–15.

Gazzola, V., & Keysers, C. (2009). The observation and execution of actions share motor and somatosensory voxels in all tested subjects: Single-subject analyses of unsmoothed fMRI data. *Cerebral Cortex, 19,* 1239–1255.

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

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

Grabski, K., Lamalle, L., Vilain, C., Schwartz, J.-L., Vallée, N., Tropres, I., et al. (2012). Functional MRI assessment of orofacial articulators: Neural correlates of lip, jaw, larynx, and tongue movements. *Human Brain Mapping, 33,* 2306–2321.

Grabski, K., Schwartz, J.-L., Lamalle, L., Vilain, C., Vallée, N., Baciu, M., et al. (2013). Shared and distinct neural correlates of vowel perception and production. *Journal of Neurolinguistics, 26,* 384–408.

Grillner, S., Robertson, B., & Stephenson-Jones, M. (2013). The evolutionary origin of the vertebrate basal ganglia and its role in action selection. *Journal of Physiology, 591,* 5425–5431.

Guenther, F. H., & Vladusich, T. (2012). A neural theory of speech acquisition and production. *Journal of Neurolinguistics, 25,* 408–422.

Haesler, S., Rochefort, C., Georgi, B., Licznerski, P., Osten, P., & Scharff, C. (2007). Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. *PLoS Biology, 5,* e321.

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A., Summerfield, A. Q., Elliott, M. R., et al. (1999). "Sparse" temporal sampling in auditory fMRI. *Human Brain Mapping, 7,* 213–223.

Hanuschkin, A., Ganguli, S., & Hahnloser, R. H. R. (2013). A Hebbian learning rule gives rise to mirror neurons and links them to control theoretic inverse models. *Frontiers in Neural Circuits, 7,* 106.

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

Heiser, M., Iacoboni, M., Maeda, F., Marcus, J., & Mazziotta, J. C. (2003). The essential role of Broca's area in imitation. *European Journal of Neuroscience, 17,* 1123–1128.

Hessler, N. A., & Doupe, A. J. (1999). Social context modulates singing-related neural activity in the songbird forebrain. *Nature Neuroscience, 2,* 209–211.

Hickok, G., Houde, J. F., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron, 69,* 407–422.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8,* 393–402.

Hope, T. M. H., Prejawa, S., Jones, O. P., Oberhuber, M., Seghier, M. L., Green, D. W., et al. (2014). Dissecting the functional anatomy of auditory word repetition. *Frontiers in Human Neuroscience, 8,* 1–17.

Houde, J. F., & Nagarajan, S. S. (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience, 5,* 82.

Hurst, J. A., Baraitser, M., Auger, E., Graham, F., & Norell, S. (1990). An extended family with a dominantly inherited speech disorder. *Developmental Medicine and Child Neurology, 32,* 352–355.

Hutchins, S., & Moreno, S. (2013). The linked dual representation model of vocal perception and production. *Frontiers in Psychology, 4,* 825.

Iacoboni, M., Woods, R., Brass, M., Bekkering, H., Mazziotta, J., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science, 286,* 2526–2528.

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

James, W. (1890). *The principles of psychology* (*Vol. 1*). New York: Holt.

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

Jarvis, E., Güntürkün, O., Bruce, L., Csillag, A., Karten, H., Kuenzel, W., et al. (2005). Avian brains and a new understanding of vertebrate brain evolution. *Nature Reviews Neuroscience, 6,* 151–159.

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

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

Kao, M. H., & Brainard, M. S. (2006). Lesions of an avian basal ganglia circuit prevent context-dependent changes to song variability. *Journal of Neurophysiology, 96,* 1441–1455.

Kao, M. H., Doupe, A. J., & Brainard, M. S. (2005). Contributions of an avian basal ganglia–forebrain circuit to real-time modulation of song. *Nature, 433,* 638–643.

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology, 9,* 718–727.

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

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

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

Kotz, S. A., Schwartze, M., & Schmidt-Kassow, M. (2009). Non-motor basal ganglia functions: A review and proposal for a model of sensory predictability in auditory language perception. *Cortex, 45,* 982–990.

Kubikova, L., Bosikova, E., Cvikova, M., Lukacova, K., Scharff, C., & Jarvis, E. D. (2014). Basal ganglia function, stuttering, sequencing, and repair in adult songbirds. *Scientific Reports, 4,* 6590.

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation. *Journal of the Acoustical Society of America, 100,* 2425–2438.

Kunzle, H. (1975). Bilateral projections from precentral motor cortex to the putamen and other parts of the basal ganglia. An autoradiographic study in *Macaca fascicularis*. *Brain Research, 88,* 195–209.

Kuypers, H. G. J. M. (1958a). Corticobulbar connexions to the pons and lower brain-stem in man. *Brain, 81,* 364–388.

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

Lai, C. S., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F., & Monaco, A. P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature, 413,* 519–523.

Leh, S. E., Ptito, A., Chakravarty, M. M., & Strafella, A. P. (2007). Fronto-striatal connections in the human brain: A probabilistic diffusion tractography study. *Neuroscience Letters, 419,* 113–118.

Lehéricy, S., Ducros, M., Krainik, A., Francois, C., Van De Moortele, P. F., Ugurbil, K., et al. (2004). 3-D diffusion tensor axonal tracking shows distinct SMA and pre-SMA projections to the human striatum. *Cerebral Cortex, 14,* 1302–1309.

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

Liégeois, F., Morgan, A. T., Connelly, A., & Vargha-Khadem, F. (2011). Endophenotypes of FOXP2: Dysfunction within the human articulatory network. *European Journal of Paediatric Neurology, 15,* 283–288.

Liu, W., & Nottebohm, F. (2005). Variable rate of singing and variable song duration are associated with high immediate early gene expression in two anterior forebrain song nuclei. *Proceedings of the National Academy of Sciences, 102,* 10724–10729.

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

Loui, P. (2015). A dual-stream neuroanatomy of singing. *Music Perception, 32,* 232–241.

Loui, P., Alsop, D., & Schlaug, G. (2009). Tone deafness: A new disconnection syndrome? *Journal of Neuroscience, 29,* 10215–10220.

Mashal, N., Solodkin, A., Dick, A. S., Elinor Chen, E., & Small, S. L. (2012). A network model of observation and imitation of speech. *Frontiers in Psychology, 3,* 1–12.

Mercado, E., III, Mantell, J. T., & Pfordresher, P. Q. (2014). Imitating sounds: A cognitive approach to understanding vocal imitation. *Comparative Cognition & Behavior Reviews, 9,* 1–57.

Molenberghs, P., Cunnington, R., & Mattingley, J. B. (2009). Is the mirror neuron system involved in imitation? A short review and meta-analysis. *Neuroscience and Biobehavioral Reviews, 33,* 975–980.

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

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

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

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

Papousek, M. (1996). Intuitive parenting: A hidden source of musical stimulation in infancy. In I. Deliège & J. Sloboda (Eds.), *Musical beginnings: Origins and development of musical competence* (pp. 88–112). Oxford: Oxford University Press.

Peeva, M. G., Guenther, F. H., Tourville, J. A., Nieto-Castanon, A., Anton, J. L., Nazarian, B., et al. (2010). Distinct representations of phonemes, syllables, and supra-syllabic sequences in the speech production network. *Neuroimage, 50,* 626–638.

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

Pfenning, A. R., Hara, E., Whitney, O., Rivas, M. V., Wang, R., Roulhac, P. L., et al. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science, 346,* 1256846.

Pfordresher, P. Q., & Brown, S. (2007). Poor-pitch singing in the absence of "tone deafness." *Music Perception, 25,* 95–115.

Pfordresher, P. Q., Brown, S., Meier, K. M., Belyk, M., & Liotti, M. (2010). Imprecise singing is widespread. *Journal of the Acoustical Society of America, 128,* 2182–2190.

Pfordresher, P. Q., Demorest, S. M., Dalla Bella, S., Hutchins, S., Loui, P., Rutkowski, J., et al. (2015). Theoretical perspectives on singing accuracy: An introduction to the special issue on singing accuracy (Part 1). *Music Perception, 32,* 227–231.

Pfordresher, P. Q., & Mantell, J. T. (2014). Singing with yourself: Evidence for an inverse modeling account of poor-pitch singing. *Cognitive Psychology, 70,* 31–57.

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

Poulson, C. L., Kymissis, E., Reeve, K. F., Andreators, M., & Reeve, L. (1991). Generalized vocal imitation in infants. *Journal of Experimental Child Psychology, 51,* 267–279.

Prather, J. F., Peters, S., Nowicki, S., & Mooney, R. (2008). Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature, 451,* 305–310.

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

Rauschecker, A. M., Pringle, A., & Watkins, K. E. (2008). Changes in neural activity associated with learning to articulate novel auditory pseudowords by covert repetition. *Human Brain Mapping, 29,* 1231–1242.

Rauschecker, J. P. (2014). Is there a tape recorder in your head? How the brain stores and retrieves musical melodies. *Frontiers in Systems Neuroscience, 8,* 149.

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience, 12,* 718–724.

Reiterer, S., Erb, M., Grodd, W., & Wildgruber, D. (2007). Cerebral processing of timbre and loudness: fMRI evidence for a contribution of Broca's area to basic auditory discrimination. *Brain Imaging and Behavior, 2,* 1–10.

Reiterer, S. M., Erb, M., Droll, C. D., Anders, S., Ethofer, T., Grodd, W., et al. (2005). Impact of task difficulty on lateralization of pitch and duration discrimination. *NeuroReport, 16,* 239–242.

Reiterer, S. M., Hu, X., Erb, M., Rota, G., Nardo, D., Grodd, W., et al. (2011). Individual differences in audio–vocal speech imitation aptitude in late bilinguals: Functional neuro-imaging and brain morphology. *Frontiers in Psychology, 2,* 1–12.

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

Rilling, J. K., Glasser, M. F., Preuss, T. M., Ma, X., Zhao, T., Hu, X., et al. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nature Neuroscience, 11,* 426–428.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3,* 131–141.

Sanvito, S., Galimberti, F., & Miller, E. H. (2007). Observational evidences of vocal learning in southern elephant seals: A longitudinal study. *Ethology, 113,* 137–146.

Segawa, J. A., Tourville, J. A., Beal, D. S., & Guenther, F. H. (2015). The neural correlates of speech motor sequence learning. *Journal of Cognitive Neuroscience, 27,* 819–831.

Shin, Y. K., Proctor, R. W., & Capaldi, E. J. (2010). A review of contemporary ideomotor theory. *Psychological Bulletin, 136,* 943–974.

Shmuelof, L., & Krakauer, J. W. (2011). Are we ready for a natural history of motor learning? *Neuron, 72,* 469–476.

Shriberg, L. D., Ballard, K. J., Tomblin, J. B., Duffy, J. R., Odell, K. H., & Williams, C. A. (2006). Speech, prosody, and voice characteristics of a mother and daughter with a 7;13 translocation affecting FOXP2. *Journal of Speech, Language, and Hearing Research, 49,* 500–525.

Simmonds, A. J., Leech, R., Iverson, P., & Wise, R. J. S. (2014). The response of the anterior striatum during adult human vocal learning. *Journal of Neurophysiology, 112,* 792–801.

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

Sohrabji, F., Nordeen, E. J., & Nordeen, K. W. (1990). Selective impairment of song learning following lesions of a forebrain nucleus in the juvenile zebra finch. *Behavioral and Neural Biology, 53,* 51–63.

Stoeger, A. S., Mietchen, D., Oh, S., de Silva, S., Herbst, C. T., Kwon, S., et al. (2012). An Asian elephant imitates human speech. *Current Biology, 22,* 2144–2148.

Studdert-Kennedy, M. (2000). Imitation and the emergence of segments. *Phonetica, 57,* 275–283.

Talairach, J., & Tournox, P. (1988). *Co-planar stereotaxix atlas of the human brain. 3-dimensional proportional system: An approach to cerebral imaging*. New York: Gerg Thieme Verlag.

Teramitsu, I., & White, S. A. (2006). FoxP2 regulation during undirected singing in adult songbirds. *Journal of Neuroscience, 26,* 7390–7394.

Trehub, S. E. (2001). Musical predispositions in infancy. In R. J. Zatorre & I. Peretz (Eds.), *The biological foundations of music* (pp. 1–16). New York: New York Academy of Sciences.

Warren, J. E., Wise, R. J. S., & Warren, J. D. (2005). Sounds do-able: Auditory-motor transformations and the posterior temporal plane. *Trends in Neurosciences, 28,* 636–643.

Watkins, K. E., Dronkers, N. F., & Vargha-Khadem, F. (2002). Behavioural analysis of an inherited speech and language disorder: Comparison with acquired aphasia. *Brain, 125,* 452–464.