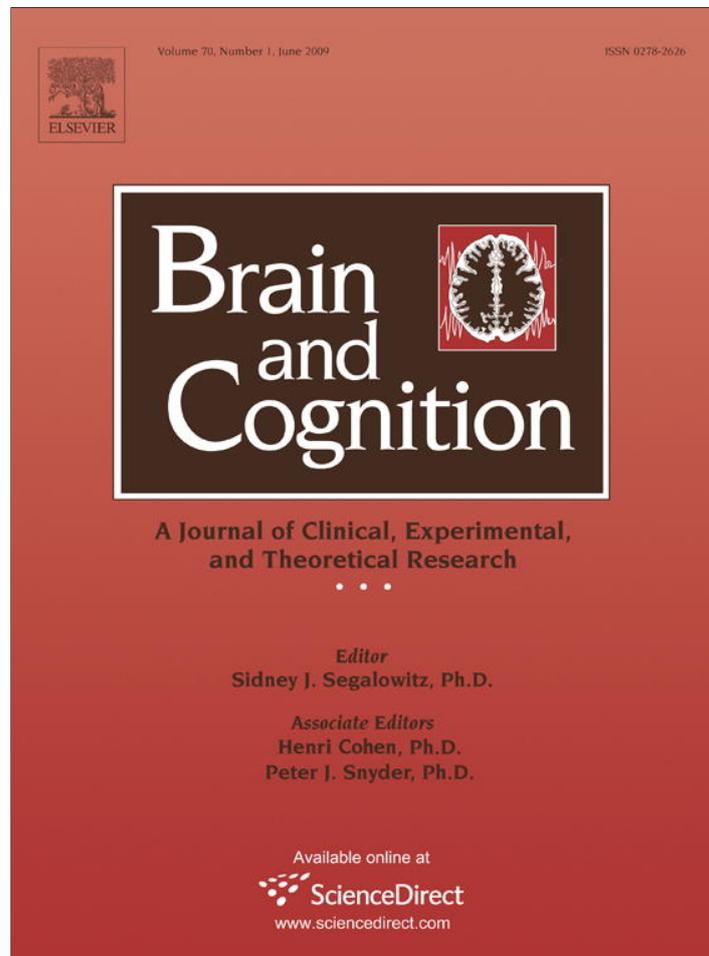


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

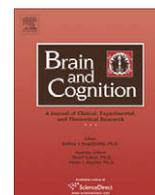
Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>

Contents lists available at [ScienceDirect](http://www.sciencedirect.com)

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c

The somatotopy of speech: Phonation and articulation in the human motor cortex

Steven Brown^{a,*}, Angela R. Laird^b, Peter Q. Pfordresher^c, Sarah M. Thelen^b, Peter Turkeltaub^d, Mario Liotti^a^a Department of Psychology, Simon Fraser University, Burnaby, BC, Canada^b Research Imaging Center, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA^c Department of Psychology, University at Buffalo, Buffalo, New York, NY, USA^d Department of Neurology, Hospital of the University of Pennsylvania, Philadelphia, PA, USA

ARTICLE INFO

Article history:

Accepted 5 December 2008

Available online 21 January 2009

Keywords:

Speech
Vocalization
Phonation
Larynx
Articulation
Brain
fMRI
Neuroimaging
Meta-analysis
ALE

ABSTRACT

A sizable literature on the neuroimaging of speech production has reliably shown activations in the orofacial region of the primary motor cortex. These activations have invariably been interpreted as reflecting “mouth” functioning and thus articulation. We used functional magnetic resonance imaging to compare an overt speech task with tongue movement, lip movement, and vowel phonation. The results showed that the strongest motor activation for speech was the somatotopic larynx area of the motor cortex, thus reflecting the significant contribution of phonation to speech production. In order to analyze further the phonatory component of speech, we performed a voxel-based meta-analysis of neuroimaging studies of syllable-singing (11 studies) and compared the results with a previously-published meta-analysis of oral reading (11 studies), showing again a strong overlap in the larynx motor area. Overall, these findings highlight the under-recognized presence of phonation in imaging studies of speech production, and support the role of the larynx motor cortex in mediating the “melodicity” of speech.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

Phonation is an important “umbrella” process when thinking about human vocalization, taking account of much of the segmental aspect of speech, of suprasegmental processes like intonation (Ladd, 1996) and lexical tone (Yip, 2002), and of singing (Sundberg, 1987). Modulation of the pitch and duration of voiced sounds underlies the melodic and rhythmic aspects of speech. The older literature on intonation employed the term “melodicity” to refer to the basic acoustic stream of voicing that occurs during speech production (Fónagy, 1981; Fónagy & Magdics, 1963).

Standard models of vocal production posit the existence of a vocal “source” – i.e., subglottal air pressure from the lungs producing vibration of the vocal-folds in the airstream – followed by “filtering” of the source’s sound wave by a series of articulators in the oral and nasal cavities, to ultimately select out certain resonant frequencies in that wave. While all vowels and most consonants require phonation, some consonants can be generated in a voiceless fashion. For fricatives like the /s/ sound, this can simply involve the generation of broadband noise at the larynx in the absence of periodic vocal-fold vibration. However, the majority of the

speech stream is phonated. For many languages, the proportion of a spoken sentence’s duration taken up by vowels alone is 40–50% (Ramus, Nespor, & Mehler, 1999). This does not take into account the degree of phonation that comes from voiced consonants, which would make the overall voiced component of a sentence’s duration even higher.

While phonation is a critical component of speech, neuroimaging studies have rarely recognized this point. Imaging studies of speech production reliably show activity in the ventral part of the precentral gyrus – corresponding with the somatotopic “orofacial” region of the motor and premotor cortices – and this activation has almost invariably been interpreted as reflecting articulation (e.g., Fox et al., 2001). The strong, if unspoken, assumption is that speech is first and foremost an articulatory process. Most studies that have sought to examine phonatory aspects of speech have (1) been perceptual rather than production studies (although see Barrett, Pike, & Paus, 2004), and (2) focused on suprasegmental processes like prosody or lexical tone rather than the basic speech stream. A handful of studies have tried to distinguish brain areas for articulation and phonation. For example, Murphy et al. (1997) compared vocalization of a simple phrase with silent mouthing of the phrase (to reveal phonation) and with mouth-closed vocalization of the phrase using the /a/ vowel alone (to reveal articulation). Their primary interest was in examining brain areas involved in respiration for speech. They identified a bilateral region of the sensorimotor cortex that was more active

* Corresponding author. Present address: Department of Psychology, Neuroscience and Behaviour, McMaster University, 1280 Main St. W, Hamilton, Ont., Canada L8S 4K1. Fax: +1 905 529 6225.

E-mail address: stebro@mcmaster.ca (S. Brown).

when speech breathing was involved than simple mouthing. Likewise, Terumitsu, Fujii, Suzuki, Kwee, and Nakada (2006) used independent components analysis (ICA) to contrast vocalization of a string of labial syllables with silent articulation of the string without voicing of the vowels or consonants. Their analysis revealed a bilateral region close to the classical tongue region associated with tongue movement and a left-dominant area dorsal to that involved in phonation.

Recent work from our lab has led to the characterization of a somatotopic representation of the larynx in the human motor cortex (Brown, Ngan, & Liotti, 2008). Related work from another lab has shown that this same general region contains a representation of the expiratory muscles as well (Loucks, Poletto, Simonyan, Reynolds, & Ludlow, 2007; Simonyan, Saad, Loucks, Poletto, & Ludlow, 2007). In fact, this area is very close to that which Murphy et al. (1997) associated with speech breathing. (For simplicity, we will refer to this general area as the “larynx motor cortex” in this article.) Hence, the two major components that comprise the vocal source appear to be in close proximity in the motor cortex, perhaps reflecting a unique cortical-level type of respiratory/phonatory coupling specific to human vocalization; for almost all other species, this coupling occurs in the brainstem alone (Jürgens, 2002). Given that our fMRI study showed that the larynx motor cortex was activated comparably by vocal and non-vocal laryngeal tasks (i.e., vocal-fold adduction alone), this area would seem like a good candidate for being a regulator of the melodicity of complex human vocalizations such as speaking and singing.

In order to examine the phonatory component of speech, we analyzed motor cortex activations for a speech production task in comparison to elemental control tasks for tongue movement, lip movement, and monotone vowel phonation, with the intent of looking for potential additivity. In a second study, we used activation likelihood estimation (ALE) meta-analysis to compare a previously-published meta-analysis of word production (Turkeltaub, Eden, Jones, & Zeffiro, 2002) with a new meta-analysis of simple phonation, namely syllable production. The goal of the combined analysis was to characterize the neural contribution of phonation to speech production, a point that has been absent in most previous neuroimaging analyses of speech production.

2. Materials and methods

2.1. Functional MRI

2.1.1. Subjects

Sixteen subjects (eight males, eight females), with a mean age of 28.4 years (ranging from 21 to 49 years), participated in the study after giving their informed consent (Clinical Research Ethics Board, University of British Columbia). Each individual was without neurological, psychiatric or audiological illness. Subjects were all fluent English speakers but were unselected with regard to handedness. Three of the subjects were left-handed.

2.1.2. Tasks

Subjects performed six oral tasks (one task per fMRI scan), each one according to a simple blocked design of 16 s of a resting condition and 16 s of an oral task. The task order was randomized across subjects. All tasks were performed with the eyes open. Four of the tasks are described in this study. (1) *Speech*. Subjects read passages aloud from the medieval epic poem *Beowulf* with the teeth together, and thus with no jaw movement. Subjects were trained to read the passages at a very slow pace (1–2 syllables per second) so as to make the rate more comparable with the following three comparison tasks. (2) *Monotone-phonation* using the schwa vowel. Subjects were instructed to sing a comfortable pitch of their choice

using the schwa vowel, with the teeth together but with a very small lip opening to permit oral air flow and avoid humming. Hence, articulatory changes should have been minimal within the task-blocks, as well as between the task and rest blocks. After each 4–6-note breath cycle, subjects were to take a gentle, controlled inspiration through the mouth. The recommended rate of vocalization was 1 Hz. This could be considered equivalently as a monovowel or monotone task. (3) *Lip protrusion*. Subjects were instructed to pucker their lips and then return them to a resting position, and to do so at a rate of roughly 1 Hz. They were encouraged to make a small gesture and to avoid contracting other facial muscles. (4) *Vertical tongue movement* within the mouth. Subjects were instructed to move the tip of their tongue from the floor of the mouth to the hard palate with the lips together but with the teeth just slightly separated so as to create adequate space for tongue movement. The recommended rate was 1 Hz. The results for the last two tasks are partially described in Brown et al. (2008). Subjects underwent a 30-min training session on a day prior to the scanning session in order to learn how to perform the tasks in a highly controlled manner with a minimum of head or body movement.

2.1.3. Magnetic resonance imaging

Magnetic resonance images were acquired with a Philips Achieva 3-Tesla MRI at the MRI Research Centre of the University of British Columbia in Vancouver. The subject's head was firmly secured using a custom head holder and “memory” pillow. Ear plugs were used to help block out scanner noise. Subjects performed each task as 16 s epochs of an oral task alternating with 16 s epochs of rest during the course of a 6'24" scan. During all tasks but speech, the name of the task (“Lips”) positioned above a cross-hair was projected from an LCD projector onto a screen mounted at the head of the MRI table, with an angled mirror on the head coil reflecting text from the screen into the participant's field of view. During the speech task, short passages from *Beowulf* were projected; a different passage was presented during each task epoch. During the rest periods for all tasks but speech, the word “Rest”, positioned above a cross-hair, was projected onto the screen. During the rest periods for the speech task, an abstract line drawing was projected so as to subtract out visual activations as much as possible, as pilot testing showed that the cross-hair alone did not achieve this. All stimuli were created and presented using Presentation software (Neurobehavioral Systems, Albany, CA).

Functional images sensitive to the “blood oxygen level dependent” (BOLD) signal were collected with a gradient echo sequence (TR = 2000 ms, TE = 30 ms, flip angle 90°, 36 slices, 3 mm slice thickness, 1 mm gap, matrix = 80 × 80, field of view = 240 mm, voxel size 3 mm isotropic), effectively covering the whole brain (145 mm of axial extent). A total of 192 brain volumes was acquired over 6'24" of scan time, corresponding with 12 alternations between 16 s epochs of rest and 16 s epochs of task.

2.1.4. Image analysis

Functional images were reconstructed offline, and the scan series was realigned and motion corrected using the methods in SPM2 (Wellcome Department of Cognitive Neurology, University College London, UK), as implemented in Matlab (Mathworks, Natick, MA). While subject motion was a concern for this study, analysis of the realignment parameters indicated that translation and rotation corrections did not exceed an acceptable level of 1.5 mm and 1.5°, respectively, for any of the participants. Following realignment, a mean functional image was computed for each run. The mean image was normalized to the Montreal Neurological Institute (MNI) template (Friston et al., 1995a, 1995b), and this transformation was then applied to the corresponding functional series. The normalized functional images (4 mm isotropic voxels) were

smoothed with an 8 mm (full-width-at-half-maximum) isotropic Gaussian filter. The BOLD response for each task-block was modeled as the convolution of a 16 s boxcar with a synthetic hemodynamic response function composed of two gamma functions. Beta weights associated with the modeled hemodynamic responses were computed to fit the observed BOLD-signal time course in each voxel for each subject using the general linear model, as implemented in SPM2. Each subject's data was processed using a fixed-effects analysis, corrected for multiple comparisons using family-wise error, with a threshold of $p < 0.05$ ($t > 4.99$) and no extent threshold. Contrast images for each task-versus-rest analysis for each subject were brought forward into a random effects analysis, where a significance level of $p < 0.025$ was employed ("false discovery rate" correction for multiple comparisons for the whole brain; Genovese, Lazar, & Nichols, 2002) and no extent threshold. The critical t value varied across contrasts and was: $t > 3.59$ for speech, $t > 4.36$ for tongue movement, $t > 4.47$ for lip movement, and $t > 5.07$ for monotone-phonation. MNI coordinates were converted into the coordinates of Talairach and Tournoux (1988) using a non-linear transformation, as implemented in the WFU PickAtlas (Maldjian, Laurienti, Kraft, & Burdette, 2003) and based on the method of Brett (imaging.mrc-cbu.cam.ac.uk/imaging/MniTalairach), except for the case of the cerebellum, where MNI coordinates are retained. This was because of errors incurred by coordinate conversion.

2.2. ALE meta-analysis

2.2.1. Inclusion criteria for papers

Meta-analysis of 11 published studies of syllable-singing was performed using activation likelihood estimation (ALE) analysis. The studies are listed in Table 2. Our inclusion criteria were: (1) that the papers provided either Talairach or MNI coordinates for their activation foci (hence excluding Özdemiir, Nortona, & Schlaug, 2006); (2) that all of the brain was imaged; (3) that only syllables were sung, but no words or sentences (hence excluding Jeffries, Fritz, & Braun, 2003, and Kleber, Birbaumer, Veit, Trevorrow, & Lotze, 2007); and (4) that overt phonation was used as part of the task (hence excluding all studies of covert production). (5) We decided to exclude articles that only presented high-level contrasts, i.e., no contrast to a low-level control condition such as rest or a perceptual baseline, as we wanted to place our focus on motor activations. On this basis, we excluded the article of Saito, Ishii, Yagi, Tatsumi, and Mizusawa (2006).

2.2.2. Activation likelihood estimation (ALE) analysis

Coordinates for activation foci from conditional contrasts were taken from the original publications. No deactivations were examined in the meta-analysis, as none of the papers reported them. We used the implementation of ALE (Laird et al., 2005a) that is contained within the BrainMap database (<http://brainmap.org>; Fox & Lancaster, 2002; Laird et al., 2005b). MNI coordinates were automatically converted to Talairach coordinates using the method of Brett cited above. All coordinates were then blurred with a full-width-at-half-maximum of 12 mm. The ALE statistic was computed for every voxel in the brain according to the algorithm developed by Turkeltaub et al. (2002). A permutation test using 5000 permutations was performed to determine the statistical significance of the ALE results, which were thresholded at $p < 0.05$ using the "false discovery rate" correction for multiple comparisons (Laird et al., 2005a). The ALE maps presented in Fig. 2 are shown overlaid onto an anatomical template generated by spatially normalizing the International Consortium for Brain Mapping (ICBM) template to Talairach space (Kochunov et al., 2002).

Fig. 2 also shows the results of a re-analysis of the Turkeltaub et al. (2002) reading data that we use as a comparison for the re-

sults of the syllable-singing meta-analysis. The data are different in two respects compared to the original publication. First, our analysis used a false discovery rate threshold of 0.05 based on 5000 permutations in order to correct for multiple comparisons, whereas the original analysis was uncorrected and used a threshold of $p < 0.0001$ based on 1000 permutations; and second, MNI coordinates were presented in the Turkeltaub analysis, whereas we converted MNI coordinates into Talairach space. In addition, Table 3 presents submaxima for several of the major peaks, some of which are not reported in the original publication. In order to make the syllable-singing and reading analyses more comparable, we applied an extent threshold of 400 mm³ to the reading analysis, corresponding to the smallest cluster reported for the syllable-singing analysis. Finally, although a new transformation procedure for converting MNI coordinates to Talairach coordinates was published (Lancaster et al., 2007) and became implemented into the ALE procedure, we chose to use Brett procedure in this analysis for two reasons. First, we wanted the fMRI results to correspond with previously-published results from this data set (Brown et al., 2008), which used the Brett transform. Second, we wanted the reading meta-analysis coordinates to match as closely as possible the published coordinates in Turkeltaub et al. (2002).

3. Results

3.1. fMRI

An analysis of the speech task vs. rest (Fig. 1 and Table 1) showed bilateral activations in the part of the motor/premotor cortex that Brown et al. (2008) identified as the larynx representation, showing ventromedial peaks (slice at z of 32) and dorsolateral peaks (slice at z of 40). A second major activation focus in the motor cortex was found in the Rolandic operculum, which we showed previously contains, at least in part, the ventral portion of the somatotopic tongue representation (Brown et al., 2008), thus reflecting articulation. Examination of the peaks at z slice 32 shows that there is much activity smeared lateral to the larynx peak. This most likely represents the labial contribution to speaking, although SPM did not identify a separate focus of activation here. Additional motor activations were seen in the supplementary motor area (SMA) and two distinct regions of the cerebellum bilaterally, namely lobules VI and VIIIA. Auditory activations were seen bilaterally in both the anterior and posterior parts of the superior temporal gyrus (STG) and sulcus, including those involved in voice perception (Belin, Zatorre, LaFalce, Ahead, & Pike, 2000). Most of the sensorimotor activations for speech were bilateral except for a left-hemisphere focus in area Spt in the posterior part of the STG.

The bottom part of Fig. 1 shows direct subtractions of tongue movement, lip movement, or monotone-phonation from the speech task, with an emphasis on the motor cortex. Subtraction of either tongue or lip movement from speech revealed a residual peak in the ventromedial larynx area, suggestive of the role of this area in phonation. Subtraction of tongue movement, but not lip movement or phonation, eliminated activity in the Rolandic operculum, suggestive of a primary role of this region in tongue movement rather than phonation. The phonation condition was the least effective subtraction control, as it failed to appreciably subtract out larynx activity from the speech condition. The potential reasons for this are discussed below. Overall, these results show that individual components of speech can be eliminated using a subtractive approach, hence arguing for a basic additivity of the speech system as well as for the common recruitment of motor-cortical regions by speech and non-speech articulator movements.

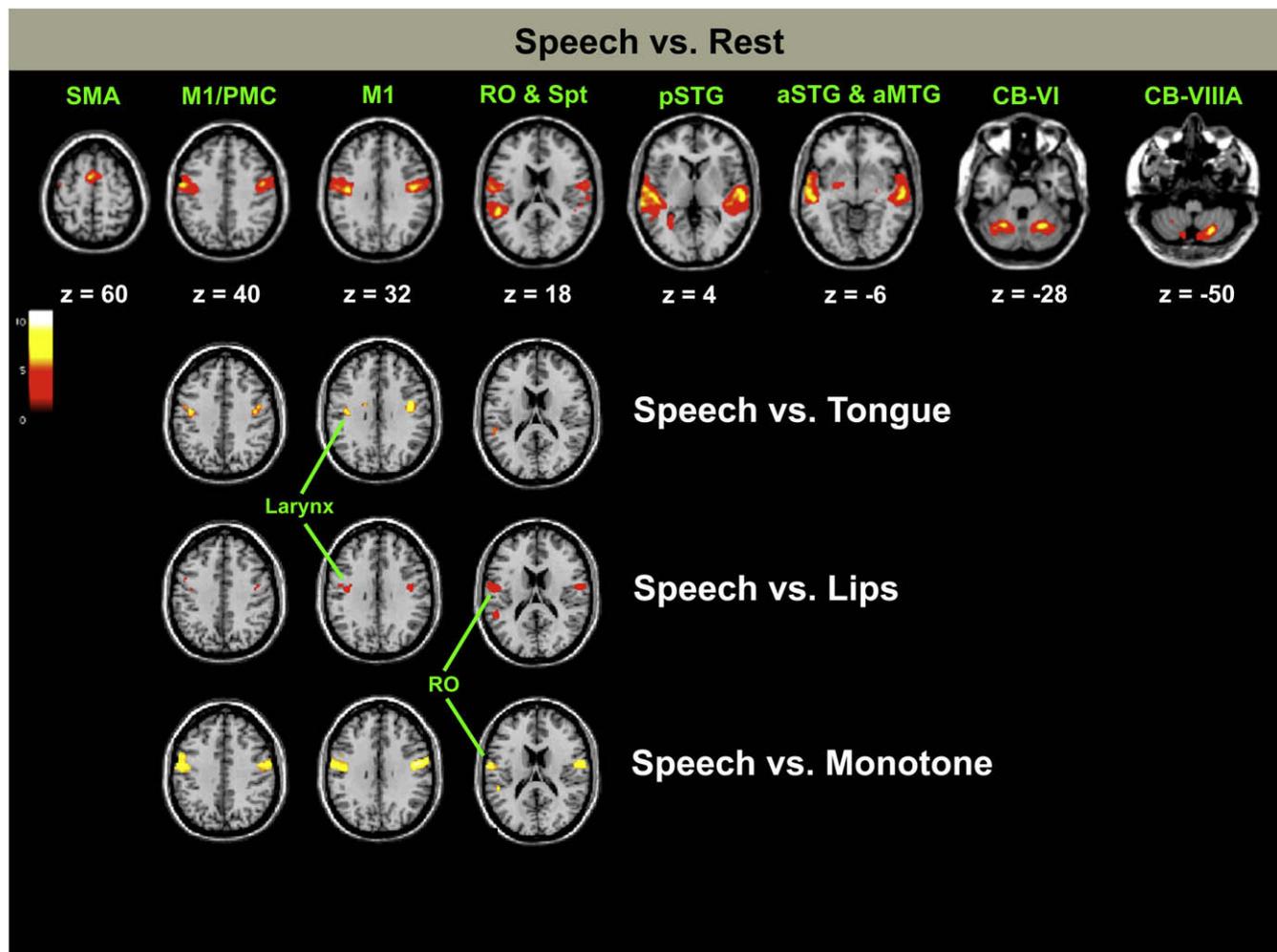


Fig. 1. fMRI activations for speech. Group activations for speech vs. rest as registered onto the MNI template brain; MNI z coordinates are given below each slice. These images are thresholded to $p < 0.025$, corrected for multiple comparisons using the false discovery rate ($t > 3.59$). The right side of the slice is the right side of the brain. Note that the coordinates listed in Table 1 are converted into Talairach coordinates, and so there are some discrepancies between the MNI slices shown for the peak activations in this figure and the Talairach coordinates listed in the table. Below this are the results of three sets of subtraction analyses, with a focus on motor cortex activations, with images thresholded to $p < 0.025$, corrected for multiple comparisons using the false discovery rate: speech vs. tongue movement ($t > 3.86$); speech vs. lip movement ($t > 3.84$); and speech vs. monotone-phonation ($t > 4.01$). Abbreviations (from left to right): SMA, supplementary motor area; M1, primary motor cortex; PMC, premotor cortex; Spt, cortex of the dorsal Sylvian fissure at the parietal–temporal junction; pSTG, posterior part of the superior temporal gyrus; aSTG, anterior part of the superior temporal gyrus; CB-VI, lobule VI of the cerebellum; CB-VIIIA, lobule VIIIA of the cerebellum; RO, Rolandic operculum; MTG, middle temporal gyrus.

3.2. ALE meta-analysis

A second approach was taken to look at the melodicity of speech, namely a comparison between a previously-published meta-analysis of overt speech (i.e., oral reading of word lists) and our own meta-analysis of 11 studies of syllable-singing. A total of 283 foci from these studies were used in the ALE meta-analysis. In contrast to the minimalist fMRI phonation task used here, almost all the phonatory tasks previously-published had a definite articulatory component to them, using syllables like /da/ and /pa/ (see Table 2 for details). The major ALE clusters are shown in Fig. 2a, with Talairach coordinates presented in Table 3.

The two meta-analyses showed common activations in both the larynx motor cortex, indicative of phonation, and the Rolandic operculum, indicative of articulation. It is interesting to note that, compared to the simple schwa vowel used in our fMRI monotone task, almost all of the tasks in the syllable-singing meta-analysis used syllables that involved articulatory transitions between consonants and vowels. Hence, while we did not see the Rolandic operculum in the fMRI phonation task, this area did indeed show up strongly in the meta-analysis, most likely reflecting the occur-

rence of articulation in these tasks. The meta-analysis findings bolster the fMRI results in highlighting the role of the larynx motor cortex in basic melodicity as well as in permitting a somatotopic assignment of phonation and lingual articulation to two regions of the motor cortex. Since syllable-singing reproduced much of the activation pattern of complex speech, it appears that, at the motoric level, speech is indeed a combination of its phonatory and articulatory components.

As a final step, we performed a comparison between the syllable-singing meta-analysis and a previously-published meta-analysis of 11 studies of overt word-reading (Turkeltaub et al., 2002), as shown in Fig. 2b and Table 4. Table 4 shows the strong overlap in activity between the peak coordinates of the syllable-singing and reading meta-analyses. The vast majority of the foci in the reading meta-analysis were present in the syllable-singing meta-analysis, again with a substantial overlap in the larynx motor cortex, this time having a good match to the ventromedial peak. Highly similar results were obtained when the syllable-singing meta-analysis was compared with another meta-analysis of overt reading, namely the data of Brown, Ingham, Ingham, Laird, and Fox (2005), looking at fluent control subjects in eight published studies of stuttering

Table 1

Speech contrasted with rest. fMRI coordinates for speech. Stereotaxic coordinates and *t*-score values for activations in the speech task contrasted with rest. MNI coordinates generated using SPM2 were converted to Talairach coordinates using the WFU Pickatlas, except for the cerebellum, in which case MNI coordinates have been retained, due to location errors that occur with conversion. Brain atlas coordinates are in millimeters along the left–right (*x*), anterior–posterior (*y*), and superior–inferior (*z*) axes. In parentheses after each brain region is the Brodmann area, except for the cerebellum, in which case the anatomical labels of Schmahmann, Doyon, Toga, Petrides, and Evans (2000) are used. Abbreviations: M1, primary motor cortex; Sup., superior; STS, superior temporal sulcus; Spt, cortex of the dorsal Sylvian fissure at the parietal–temporal junction. The threshold is *t* > 3.59.

Region	<i>x</i>	<i>y</i>	<i>z</i>	<i>t</i>
Frontal				
Right				
M1 larynx, ventromedial (4)	44	–10	30	8.14
M1 larynx, dorsolateral (6)	44	–6	33	8.24
Rolandic operculum (43/4)	59	–5	17	6.11
Left				
M1 larynx, ventromedial (4)	–40	–12	30	10.10
Supplementary motor area (6)	–4	7	55	8.53
M1 larynx, dorsolateral (6)	–55	–4	37	8.36
Rolandic operculum (43/4)	–50	–9	23	7.73
Premotor cortex (6)	–42	–6	39	5.87
Temporal				
Right				
Middle temporal gyrus (21)	58	–19	–1	11.00
Superior temporal gyrus (22)	52	–32	11	6.96
Superior temporal sulcus (21/22)	52	–38	9	5.82
Superior temporal gyrus (22/42)	65	–21	16	4.31
Left				
Superior temporal gyrus (22)	–57	–19	1	13.00
Posterior STG/area spt (22)	–50	–36	17	9.14
Middle temporal gyrus (21)	–57	–27	1	12.40
Middle temporal gyrus (21)	–59	–12	–6	8.37
Middle temporal gyrus (21)	–53	–48	6	3.76
Auditory assoc. cortex (42)	–57	–13	14	8.82
STS/anterior STG	–61	–2	–3	7.28
Primary auditory cortex (41)	–38	–31	7	6.56
Cerebellum (MNI coordinates)				
Right				
Lobule VIIIA	24	–66	–50	10.50
Lobule VI	20	–60	–20	9.52
Lobule VI	28	–58	–28	9.00
Vermis, lobule VIIB	8	–74	–44	7.25
Lefts				
Lobule VI	–24	–58	–28	8.58
Vermis, lobule VI	–12	–64	–20	8.09
Lobule VI/crus I	–34	–62	–28	6.81
Lobule VIIIA	–12	–72	–50	4.70
Lobule VIIIA	–24	–58	–46	3.67
Subcortical				
Putamen	–24	–6	–5	6.10
Putamen	–22	–4	8	4.00

Table 2

Studies included in the syllable-singing meta-analysis. Eleven studies were included in the ALE meta-analysis. Mod. = imaging modality. “*n*” is the number of subjects for each study. “Gen.” refers to the gender of the subjects in the study, where M is male and F is female. The fifth and sixth columns describe the vocal task and control task for each study. The last column describes the major articulatory vehicle used in each study to gauge phonation. For Bohland and Guenther, the designation /CV/ in the last column means that a series of different consonant-vowel syllables was used as stimuli in their study.

Reference	Mod.	<i>n</i>	Gen.	Vocal task	Control task	Syllable
1. Perry et al. (1999)	PET	13	M/F	Monotone singing	Auditory control task	la
2. Riecker et al. (2000a)	fMRI	18	M/F	Singing a familiar melody	Rest	?
3. Riecker et al. (2000b)	fMRI	10	M/F	Monosyllable production	Rest	ta, stra
4. Riecker, Wildgruber, Dogil, Grodd, and Ackermann (2002)	fMRI	12	M/F	Isochronous syllable production	Perceptual baseline	pa
				Rhythmic syllable production	Perceptual baseline	pa
5. Brown et al. (2004)	PET	10	M/F	Monotone singing	Rest	da
				Melody repetition	Rest	da
				Harmonization	Rest	da
6. Wilson, Saygin, Sereno, and Iacoboni (2004)	fMRI	10	M/F	Monosyllable production	Rest	pa, gi
7. Riecker et al. (2005)	fMRI	8	M/F	Monosyllable production	Listening to clicks	pa
8. Brown et al. (2006)	PET	10	M/F	Melody generation	Rest	da
9. Sörös et al. (2006)	fMRI	9	M/F	Vowel production	Rest	ah
10. Riecker et al. (2006)	fMRI	8	M/F	Monosyllable production	Listening to clicks	pa
11. Bohland and Guenther (2006)	fMRI	13	M/F	Trisyllable production	Visual baseline	/CV/

(data not shown). However, we chose to focus here on the Turkeltaub analysis because it was based on 172 foci, compared to only 73 foci for the stuttering controls. The overall profile, however, was very similar.

4. Discussion

In this study, we attempted to look at speech in a somatotopic manner, and especially to illuminate the role of phonation in speech production. We use these analyses to formulate a general model of vocalization in the brain.

4.1. Phonation and articulation

Our previous fMRI study (Brown et al., 2008) established a representation of the larynx in the motor cortex, one which overlaps an area involved in voluntary control of expiration (Loucks et al., 2007; Simonyan et al., 2007). Using this motor cortex focus as a reference, we were able to demonstrate for the first time that connected speech gives its principal motor cortex activation in the larynx area, thereby supporting the notion that much of the speech signal is voiced, including all vowels and a majority of consonants. Previous neuroimaging studies on speech production have not made this point about phonation, and have instead talked about activity in the “mouth” or “face” area of the motor cortex (e.g., Fox et al., 2001), with the implication being that speech is mainly articulatory. Knowing the location of the larynx area, we were able to interpret residual activations in the motor strip as being related to articulation, mainly in the Rolandic operculum for tongue movement and the region lateral to the larynx area for lip movement. This is a first step toward a somatotopic dissection of phonation and articulation in the cortical motor system. The study of Terumitsu et al. (2006) seemed poised to make the same point, in that the authors compared phonated vs. mouthed versions of the same polysyllable string. However, their analyses did not involve a direct contrast between the voiced and unvoiced tasks, and what they called “phonation” in their ICA analysis included articulation as well as phonation, as evidenced by ICA clusters in the Rolandic operculum.

The results with the speech task match very closely the findings of two voxel-based meta-analyses of overt reading. Turkeltaub et al. (2002) published an activation likelihood estimation (ALE) meta-analysis of 11 studies of oral reading, and found the region of greatest concordance across these studies in the motor cortex to be at $-48, -12, 36$, and $44, -10, 34$, very close to our ventromedial speech peaks at $-40, -12, 30$, and $44, -10, 30$. Likewise,

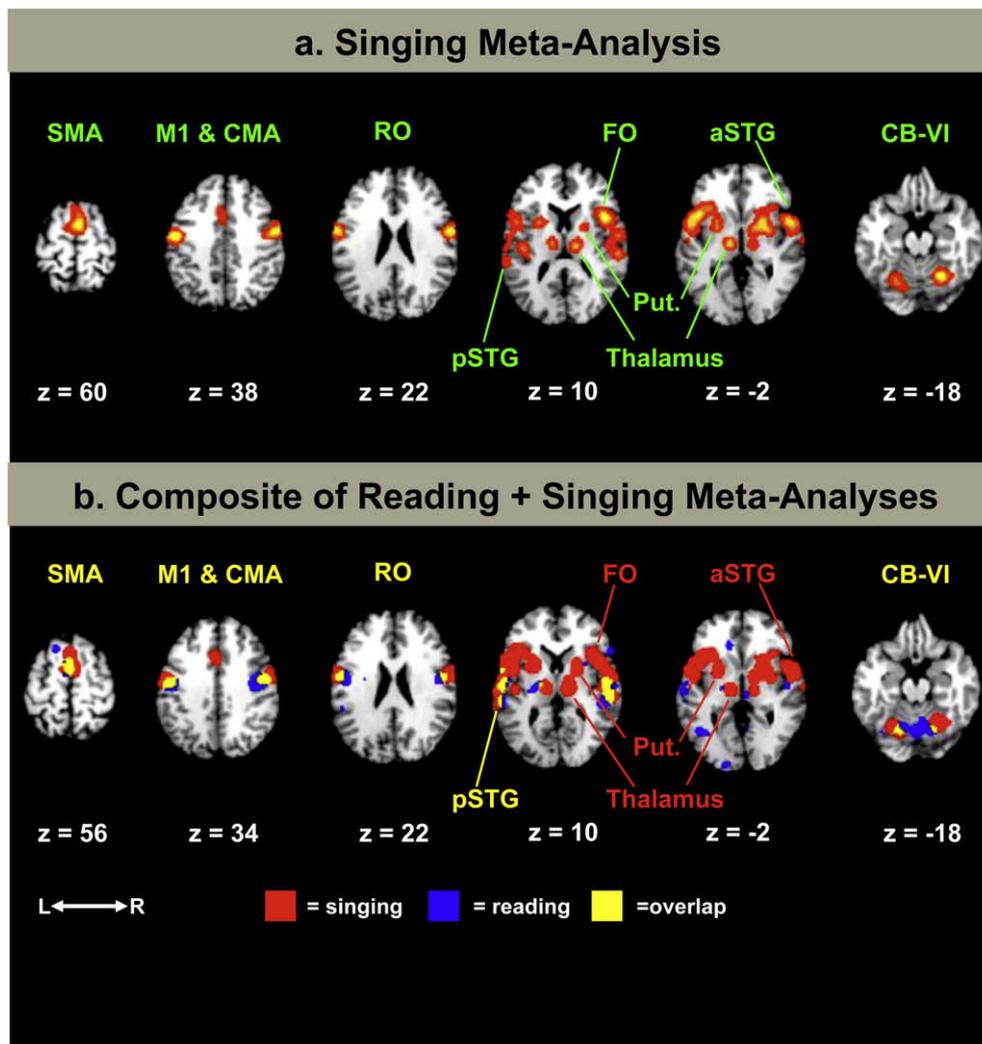


Fig. 2. Major ALE foci from the meta-analysis. (a) Major ALE foci for the syllable-singing meta-analysis. Principal sites of activation are labelled; some bilateral cortical activations are labelled on only one side of the brain. Talairach z coordinates are shown below each slice. (b) Comparison of the ALE foci from the current meta-analysis of syllable-singing and a meta-analysis of overt reading performed by Turkeltaub et al. (2002). The color scheme for ALE activations in this panel is the following: red = syllable-singing, blue = reading, yellow = overlap. The labels in this panel highlight the vocal-motor areas shown by the meta-analysis to have large cross-laboratory concordance. For the bottom panel only, yellow labels refer to common activations between speaking and syllable-singing, and red labels refer to foci unique to syllable-singing. The Talairach coordinates for the slices are shown at the bottom of the figure. In choosing slice levels for this composite analysis, we attempted to present slices that were intermediate between the peak activations for syllable-singing and those for reading where differences existed between them (see Table 4 for coordinates). The only location where this does not work well is the slice at $z = 32$, which gives the impression that the CMA is uniquely present in syllable-singing, when the speech focus is actually 7 mm higher, and thus not present in this slice. Hence, the label for the CMA is colored yellow instead of red. The right side of a slice is the right side of the brain. The threshold for both analyses is $p < 0.05$, corrected for multiple comparisons using the false discovery rate. Abbreviations (from left to right) are: SMA, supplementary motor area; M1, primary motor cortex; CMA, cingulate motor area; RO, Rolandic operculum; pSTG, posterior part of the superior temporal gyrus; FO, frontal operculum; Put., putamen; aSTG, anterior part of the superior temporal gyrus; CB-VI, lobule VI of the cerebellum. (For interpretation of the references in color in this figure legend, the reader is referred to the web version of this article.)

Brown et al. (2005) performed two parallel ALE meta-analyses of eight studies of oral reading in stutterers and fluent controls, respectively. The peak M1 activations for the control subjects were at $-49, -9, 32$, and $54, -10, 34$, and those for the stutterers were at $-45, -16, 31$, and $48, -12, 32$, again quite close to the ventromedial M1 peaks for the speech task in this study. Both of these meta-analyses identified the larynx area as a major location of activation during oral reading. They also found bilateral activations at the Rolandic operculum, very close to our fMRI tongue coordinates. Hence, the general pattern seen to emerge from imaging studies of speech production is two major sites of activation in the motor cortex: the larynx area deep in the central sulcus, and the tongue area in the Rolandic operculum. While other processes are clearly critical for speech production – not least muscular activity in the lips, velum and pharynx – larynx and tongue activities might be

the most readily identifiable ones because of their distance in the motor cortex. For example, in our previous imaging study (Brown et al., 2008), lip and tongue movement showed a region of overlapping activity, although this was dorsal to tongue-related region of the Rolandic operculum.

Given our separability of activity in the larynx area and Rolandic operculum during simple phonation and tongue movement, respectively, and their combination during speech (in addition to presumptive activity in lip-related areas), there does seem to be a basic additivity of phonation and articulation that comes into play during speech production. Looking to the subtraction analysis, we obtained mixed results. While tongue and lip movement nicely subtracted out articulation-related activity in the motor cortex during speech, the monotone-phonation task was not very effective at subtracting out the larynx peak of speech. Interestingly, a

Table 3

Syllable-singing ALE clusters. Major ALE clusters for the syllable-singing meta-analysis. The table shows the principal ALE clusters derived from the meta-analysis of the 11 studies of syllable-singing. After each anatomical name in the “region” column is the Brodmann number in parentheses. The columns labelled as x, y, and z are the Talairach coordinates for the weighted center of each cluster. The ALE score shown is the true value multiplied by 10³. The right column shows the size of each cluster in mm³. Note that the rightmost column specifies three large clusters, submaxima of which are identified in the column for cluster size (mm³). Abbreviations: M1, primary motor cortex; Ant., anterior; Sup., superior.

Region	x	y	z	ALE ($\times 10^{-3}$)	(mm ³)
<i>Frontal</i>					
<i>Right</i>					
Frontal operculum (44)	44	10	10	20.1	34,200 (Cluster 1)
Rolandic operculum (4/6/43)	60	-4	24	19.9	Cluster 1
M1 larynx (4/6)	54	-4	38	19.2	Cluster 1
Supplementary motor area (6)	6	0	60	19.2	11,040 (cluster 3)
<i>Left</i>					
M1 larynx (4/6)	-50	-10	40	21.5	28,920 (cluster 2)
Rolandic operculum (4/6/43)	-58	-2	22	18.3	Cluster 2
Cingulate motor area (32/6)	-2	6	46	14.1	Cluster 3
Cingulate motor area (32/6)	-2	16	32	10.7	Cluster 3
<i>Temporal</i>					
<i>Right</i>					
Ant. Sup. temporal gyrus (22)	52	4	-4	17.8	Cluster 1
Superior temporal gyrus (22)	58	-28	10	13.5	Cluster 1
Superior temporal gyrus (22)	58	-10	10	11.5	Cluster 1
<i>Left</i>					
Ant. sup. temporal gyrus (22)	-44	10	0	20.5	Cluster 2
Primary auditory cortex (41)	-40	-22	10	12.9	Cluster 2
Superior temporal gyrus (22)	-60	-22	4	11.8	Cluster 2
Primary auditory cortex (41)	-34	-34	14	11.2	Cluster 2
Superior temporal gyrus (22)	-58	-36	8	10.5	Cluster 2
<i>Subcortical</i>					
<i>Right</i>					
Putamen	24	2	-2	17.7	Cluster 1
Ventral thalamus	14	-20	6	17.3	Cluster 1
<i>Left</i>					
Ventral thalamus	-10	-18	0	17.0	3352
Putamen	-22	4	4	15.8	Cluster 2
<i>Cerebellum</i>					
<i>Right</i>					
Lobule VI	24	-60	-20	21.7	4056
Vermis V/VI					1672
<i>Left</i>					
Lobule VI	-34	-50	-30	15.4	1672
Lobule VI	-22	-64	-18	13.0	2376
Vermis V/VI	-4	-56	-28	11.0	504

similar result was found in the study of Murphy et al. (1997). Their contrast was better matched than ours in that they compared the vocalization of a phrase with mouth-closed vocalizing of the same phrase using the /a/ vowel. Hence, much about the melody and rhythm of the original phrase should have been contained in the unarticulated version. Their subtraction revealed bilateral peaks in the sensorimotor cortex quite close to the ventromedial larynx area. Why might the larynx activation during speech be difficult to subtract out with phonatory control tasks, especially given the efficiency of the subtraction of articulatory areas using articulatory controls tasks? One speculation is that co-articulation during speech production may activate the larynx area in a much stronger manner than tasks that involve a single articulatory posture, such as during the monovowel tasks used in this study and that of Murphy et al. (1997). Likewise, speech tasks show an oscillatory cycling between voiced and unvoiced sounds that is not seen in the controls tasks. Given the overlap in the larynx coordinates between the reading and syllable-singing meta-analyses, the effect that we and Murphy et al. are seeing is most likely quantitative rather than qualitative. Further work is needed to enlighten this point, not

Table 4

Syllable singing vs. reading meta-analyses. Meta-analysis coordinates for syllable-singing vs. reading. Talairach coordinates for the clusters for the syllable-singing meta-analysis (current study) and reading meta-analysis (from Turkeltaub et al., 2002) are shown side by side. Note the strong overlap in activation peaks for the two analyses. Abbreviations: M1, primary motor cortex; Ant., anterior; Sup., superior.

Region	Syllable singing			Reading		
	x	y	z	x	y	z
<i>Frontal</i>						
<i>Right</i>						
Frontal operculum (44)	44	10	10			
Frontal operculum (45)				56	20	4
Rolandic operculum (4/6/43)	60	-4	24	48	-3	24
M1 larynx (4/6)	54	-4	38	44	-10	34
Supplementary motor area (6)	6	0	60	0	2	54
<i>Left</i>						
M1 larynx (4/6)	-50	-10	40	-46	-12	36
Rolandic operculum (4/6/43)	-58	-2	22	-48	-8	22
Cingulate motor area (32/6)	-2	6	46	-2	12	45
Cingulate motor area (32/6)	-2	16	32			
Frontal operculum (44)				-52	4	10
<i>Temporal</i>						
<i>Right</i>						
Ant. sup. temporal gyrus (22)	52	4	-4			
Superior temporal gyrus (22)	58	-28	10	54	-28	6
Superior temporal gyrus (22)	58	-10	10	54	-14	8
<i>Left</i>						
Ant. sup. temporal gyrus (22)	-44	10	0			
Primary auditory cortex (41)	-40	-22	10			
Superior temporal gyrus (22)	-60	-22	4	-52	-14	4
Primary auditory cortex (41)	-34	-34	14			
Superior temporal gyrus (22)				-56	-43	13
Superior temporal gyrus (22)	-58	-36	8	-56	-32	4
<i>Subcortical</i>						
<i>Right</i>						
Putamen	24	2	-2			
Ventral thalamus	14	-20	6			
<i>Left</i>						
Ventral thalamus	-10	-18	0	-18	-16	2
Putamen	-22	4	4			
<i>Cerebellum</i>						
<i>Right</i>						
Lobule VI	24	-60	-20	28	-57	-21
Vermis V/VI				16	-66	-16
<i>Left</i>						
Lobule VI	-34	-50	-30	-36	-38	-24
Lobule VI	-22	-64	-18	-14	-66	-16
Vermis V/VI	-4	-56	-28			

least an analysis of potential neural sub-domains within the larynx motor cortex for vocal-fold tension vs. relaxation, and abduction vs. adduction.

4.2. A neural model of vocalization

We would like to consolidate the results of the fMRI experiment and meta-analyses into a model of vocalization (Fig. 3), one that focuses on the generation of sounds at the vocal source, and hence phonation. A very similar model of vocal production is presented by Bohland and Guenther (2006), as discussed below. The fMRI monotone-phonation task as well as the 5-note phonation task used in Brown et al. (2008) were designed to be as pure a model of phonation as possible, minimizing the contribution of articulation to the brain activations. We would like to consider the activation pattern of these tasks as a minimal model of “primary” areas for phonation, and then contrast that with data from the fMRI speech task and the two meta-analyses in order to characterize “secondary” areas that may tap more into articulation or general orofacial functioning than phonation.

The primary vocal circuit consists principally of three motor areas: (1) the larynx motor cortex and associated premotor cortex;

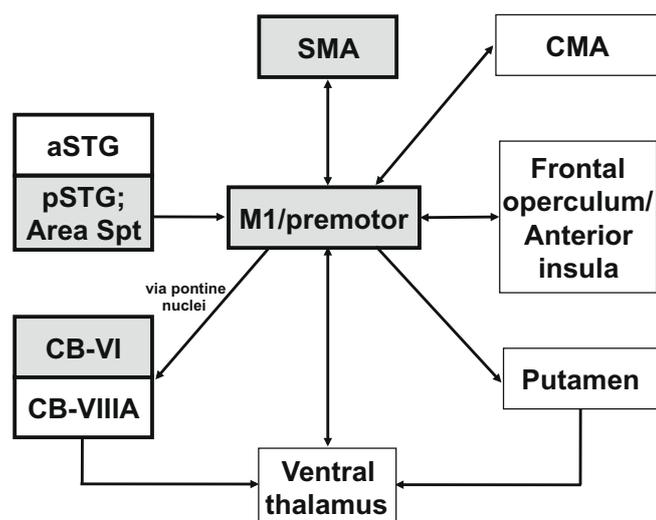


Fig. 3. Key brain areas of the vocal circuit. Shaded boxes represent “primary” areas that are principal regions for the control of phonation in speaking and singing. White boxes represent “secondary” areas that are less reliably activated during phonation and that might be more important for articulation. See text for details. This is not meant to be a comprehensive connectivity diagram. The focus is placed on the connectivity between these multiple areas and the primary motor/premotor cortex, rather than on connections among the other areas. Connectivity data is based principally on the afferent and efferent connections of the M1 larynx area of the Rhesus monkey, as described in *Simonyan and Jürgens (2002, 2003, 2005a, 2005b)*. The projection from the motor cortex to the cerebellum is via the pontine nuclei. As described in the text, lobule VIII of the cerebellum may turn out to be a primary area, but many imaging studies, especially PET studies, have not included this part of the cerebellum in their field of view. Abbreviations: SMA, supplementary motor area; CMA, cingulate motor area; pSTG, posterior part of the superior temporal gyrus; aSTG, anterior part of the superior temporal gyrus; Spt, cortex of the dorsal Sylvian fissure at the parietal–temporal junction.

(2) lobule VI of the cerebellum, and (3) the SMA. The primary auditory areas are Heschl’s gyrus and the auditory association cortex of the posterior STG, including area Spt. Secondary vocal areas include: (1) the Rolandic operculum (the ventral part of the motor cortex, hence included with the M1/premotor box in the figure), (2) the putamen and ventral thalamus, (3) cingulate motor area, and (4) frontal operculum/anterior insula. In *Fig. 3*, primary areas are shown with shaded boxes, and secondary areas with white boxes. The connectivity model in the diagram is largely based on the connections of the larynx motor cortex in the Rhesus monkey (*Simonyan & Jürgens, 2002, 2003, 2005a, 2005b*), in which most of the areas listed are reciprocally connected with the motor cortex, the exceptions being the cerebellum and putamen, which feed back to the cortex indirectly via the ventral thalamus. Regarding auditory areas, it is not known if they project directly to the motor cortex or if they have to pass through a relay like Broca’s area, as posited in the standard Geschwind model of speech (*Catania, Jones, & Fitches, 2005*). In the monkey, there is a minor projection from the posterior STG to the larynx motor cortex (*Simonyan & Jürgens, 2002, 2005b*); hence, such a pathway could exist in humans as well. Preliminary diffusion tensor imaging work from our lab suggests that there is indeed direct connectivity between temporo-parietal auditory areas and the orofacial precentral gyrus via the arcuate fasciculus (unpublished observations).

Lobule VI of the posterior cerebellum showed the highest ALE score of any brain region in the meta-analysis. Somatotopic analysis has demonstrated that this is indeed an orofacial part of the cerebellum (*Grodd, Hülsmann, Lotze, Wildgruber, & Erb, 2001*). We showed that this region is activated by lip movement and tongue movement as well as vocalization. Hence, while this region seems to be an obligatory component of the vocal circuit, there is proba-

bly little about it that is voice-specific, although there may be somatotopic sub-domains for each effector within this general area. This stands in contrast to lobule VIIIA of the ventral cerebellum, which was activated by both speech and monotone-phonation but which did not show activity for lip and tongue movement (although see *Watanabe et al. (2004)* for activity in this region during tongue movement) or show ALE foci in either meta-analysis. Given that half of the studies in the syllable-singing meta-analysis were PET studies, and given the fact that many older PET machines had an axial span of only 10 cm, it is likely that the ventral part of the cerebellum was cut off in many of the studies used in the meta-analysis (e.g., *Brown, Martinez, Hodges, Fox, & Parsons, 2004; Brown, Martinez, & Parsons, 2006*; see *Petacchi, Laird, Fox, & Bower, 2005*, for a discussion of this topic). Hence, lobule VIIIA may be a brain area that has been under-represented in studies of overt vocalization thus far and may therefore be expected to appear with greater frequency in future publications of speech and song (e.g., *Bohland & Guenther, 2006; Riecker et al., 2005*).

The SMA is one of a handful of brain areas which when lesioned can give rise to mutism, and stimulation of this area can elicit vocalization in humans but not monkeys (*Jürgens, 2002*). The SMA is organized somatotopically (*Fontaine, Capelle, & Duffau, 2002*) but, as with lobule VI of the cerebellum, there is no information as to whether there are effector-specific zones within the somatotopic orofacial area of the SMA. While the SMA is routinely activated in studies of both speech and song production, its exact role is unclear. The SMA is classically associated with activities like bimanual coordination (*Carson, 2005*), and stimulation of the SMA can lead to simultaneous activation of linked effectors, such as the whole arm. It is thus reasonable to presume that the SMA plays some role in the sequential coordination of effectors during vocal production, although this area is clearly activated when single effectors such as the lips or tongue are used. In *Indefrey & Level’s (2004)* qualitative meta-analysis of 82 studies of single-word processing, they argued that the SMA was involved in articulatory planning. In addition, they found that the SMA was active in both covert and overt word-reading tasks. In support of this, the SMA has also been found to be active in many studies of covert singing (*Callan et al., 2006; Halpern & Zatorre, 1999; Riecker, Ackermann, Wildgruber, Dogil, & Grodd, 2000a*). Hence, the SMA plays some role in motor planning, motor sequencing, and/or sensorimotor integration, but the exact role in vocalization is not well understood.

The putamen gave one of the most complex profiles of any area in these analyses. No activity was seen for the fMRI monotone-phonation task, whereas there was a strong left-hemisphere focus for speech. That said, the putamen showed very strong ALE foci bilaterally in the syllable-singing meta-analysis and reasonably good concordance across its contributing studies, hence creating an inconsistency between the fMRI study and the meta-analysis. One potential resolution to this inconsistency is to posit that the putamen is more important for articulation than phonation. In the fMRI study, we found more putamen activity for lip movement and tongue movement than for simple phonation. Likewise, many studies have shown activity in the putamen during lip movement, tongue movement, and voluntary swallowing (*Corfield et al., 1999; Gerardin et al., 2003; Martin et al., 2004; Rotte, Kanowski, & Heinze, 2002; Watanabe et al., 2004*). One problem with this interpretation is that damage to the basal ganglia circuit gives rise to severe dysphonia in addition to articulatory problems (*Merati et al., 2005*). This would seem to suggest that the basal ganglia play a direct role in phonation. It is interesting in this regard that the only major voice therapy that seems successful at ameliorating Parkinsonian dysphonia, namely Lee Silverman Voice Therapy (*Ramig, Countryman, Thompson, & Horii, 1995; Ramig et al., 2001*), is a phonation-based therapy that indirectly improves articulation as

a by-product (Dromey, Ramig, & Johnson, 1995; Sapir, Spielman, Ramig, Story, & Fox, 2007). The role of the putamen in phonation and articulation is in need of further exploration. For the time being, we put it in the category of “secondary” areas. We do the same for the ventral thalamus. Its co-occurrence with the putamen (i.e., both were absent in the fMRI monotone task, both were present in the syllable-singing meta-analysis, and the thalamus was only present in articles that reported putamen activation in the syllable-singing meta-analysis) probably reflects the connectivity of the basal ganglia, which sends its output from the internal segment of the globus pallidus to anterior parts of the ventral thalamus. The cerebellum’s projection to the cerebral cortex also passes through a part of the ventral thalamus (posterior to the basal ganglia projection), and so it is unclear why there should be an absence of ventral thalamus activation in the presence of strong cerebellar activity. The thalamus showed relatively low concordance across studies in the meta-analysis.

One interesting point of reference with regard to the basal ganglia comes from the studies of Riecker et al. (2005), Riecker, Kassubek, Groschel, Grodd, and Ackermann (2006), which were included in the meta-analysis. These studies examined the tempo of vocalization, looking at monosyllable /pa/ repetitions over the range of 2–6 Hz. What was found was that activity in lobules VI and VIIIA of the cerebellum showed positive correlations with syllable rate whereas activity in the putamen and caudate nucleus showed negative correlations. Putamen activity decreased monotonically for speaking rates ranging from 2 to 6 Hz (Riecker, Kassubek, Groschel, Grodd, & Ackermann, 2006); the two cerebellar regions showed the reverse pattern. Our profile of high cerebellum and low putamen does not follow from the assumption that these patterns would extend to 1 Hz, the suggested production rate for our fMRI monotone task. Again, the absence of articulatory changes in our singing task may be a more important factor than tempo *per se* in explaining the absence of putamen activity.

The cingulate motor area gave low concordance in the meta-analysis, and was not found to be active in the speech or monotone-phonation fMRI tasks. Unlike the larynx motor cortex, the CMA is the only cortical part of the monkey brain which, when lesioned, disrupts vocalization (Sutton, Larson, & Lindeman, 1974, but see Kirzinger & Jürgens, 1982). The projection from the cingulate cortex to the periaqueductal gray is thought to represent an ancestral vocalization pathway in primates that is perhaps more important for involuntary vocalizations than voluntary ones like speech. This area may indeed be more involved in emotive vocalizations than learned vocalizations such as speech and song in humans. It is interesting to note that almost all of the studies in the meta-analysis that showed CMA activation employed monotone tasks rather than melodic singing tasks. Hence, the CMA may have some preference for simple vocal tasks, as shown by its activation in monotone (Brown et al., 2004; Perry et al., 1999), monovowel (Sörös et al., 2006), and monosyllable (Bohland & Guenther, 2006; Riecker et al., 2006) tasks. This hypothesis is consistent with the reading study of Barrett et al. (2004), in which subjects had to read semantically-neutral passages under conditions of either happy or sad mood induction. Regressions with affect-induced pitch range showed that the more monotonous the speech became during sad speech, the greater the activity in the CMA. The major Talairach coordinate for this regression was at $-8, 18, 34$, which corresponds quite well with one of the CMA coordinates from the syllable-singing meta-analysis at $-2, 16, 32$. CMA activity may thus be sensitive to melodic complexity, showing a preference for low-complexity vocal tasks having minimal pitch variation, which may reflect its evolution from a system involved in simple, stereotyped vocalizations. Might the CMA be the brain’s “chant” center? Further work is needed to clarify the role of the cingulate cortex in vocalization.

The frontal operculum and medial-adjacent anterior insula represent yet another difficult case for our model. As with the putamen, activity in this region was much stronger in the meta-analysis than the fMRI monotone task. We again make the speculation that this area encodes generalized orofacial functions and thus might be equally involved in articulation and phonation. The fMRI study showed comparable activity in the frontal operculum for lip movement and tongue movement as for vocalization. This casts doubts on a phonation-specific role of this region. In addition, the most typical type of symptom associated with damage to the anterior insula is apraxia of speech and not dysphonia alone (Jordan & Hillis, 2006; Ogar, Slama, Dronkers, Amici, & Gorno-Tempini, 2005). Hence, damage to this region is much more likely to result in articulatory deficits than phonatory ones, although both seem to co-occur. As Ogar et al. (2005) point out: “Prosodic deficits, however, are thought to be a secondary effect of poor articulation” (p. 428). It is for these reasons that we put the frontal operculum and adjacent anterior insula into the category of “secondary” areas for vocalization. Several models of vocal production have ascribed an important role for the anterior insula in phonological processing (Ackermann & Riecker, 2004; Bohland & Guenther, 2006; Indefrey & Level, 2004; Riecker et al., 2005, 2006). In Indefrey and Level’s (2004) meta-analysis, they associated the anterior insula most strongly with “phonological code retrieval”, which is a process of searching for phonological words that match a lexically selected item. They found less evidence for a role of the anterior insula in actual speech production, a result counter to the perspective of Ackermann and Riecker (2004). Riecker et al. (2006) found that activity in the insula increased monotonically with syllable rate, hence showing a similar profile to the cerebellum (as well as larynx motor cortex and SMA). So the frontal operculum/anterior insula is almost certainly a vocal-motor area, but its exact role is in need of further analysis.

The model in Fig. 3 shows striking similarities with the “basic speech production network” proposed by Bohland and Guenther (2006), which includes all of the areas mentioned here. In fact, there is no region of disagreement between our model and theirs. Perhaps the only motivational difference relates to our goal of defining a network of vocal production based on phonation, leading to our distinction between primary and secondary areas for vocalization. Their model was based on a series of syllable tasks, ranging from simple to complex trisyllables. Hence, articulation was an important component of all of their tasks. It is possible that a task based on vowels alone would yield different results. For example, the vowel production task of Perry et al. (1999) failed to show activity in some of the areas that we have speculated to be associated with articulation (e.g., putamen) but did show activity in others (Rolandic operculum, frontal operculum), whereas the vowel production task of Sörös et al. (2006) failed to show activity in the Rolandic operculum but did show activity in the putamen and frontal operculum. Further work is clearly needed to verify the phonation network postulated in our primary areas.

5. Conclusions

Using two complementary comparisons between speech and non-speech oral tasks (fMRI and meta-analysis), we have attempted to disentangle phonation and articulation in speech, and have shown that motor-control models like the “source-filter” model can be represented somatotopically in the motor cortex. A principal site of activation for speech is the larynx representation in the motor cortex, in keeping with the overwhelmingly voiced nature of speech. Additional activity in the Rolandic operculum for tongue movement and other parts of the motor cortex

contribute to an overall sense of additivity of phonation and articulation during speech production.

Acknowledgments

This work was supported by a grant to SB from the Grammy Foundation. ARL and SMT were supported by the Human Brain Project of the NIMH (R01-MH074457-01A1), and PQP by NSF grant 0642592. We thank Trudy Harris, Jennifer McCord, and Burkhard Mädler at the MRI Research Centre of the University of British Columbia for expert technical assistance. We thank Roger Ingham (University of California at Santa Barbara) for critical reading of a previous version of the manuscript.

References

- Ackermann, H., & Riecker, A. (2004). The contribution of the insula to motor aspects of speech production: A review and a hypothesis. *Brain and Language*, *89*, 320–328.
- Barrett, J., Pike, G. B., & Paus, T. (2004). The role of the anterior cingulate cortex in pitch variation during sad affect. *European Journal of Neuroscience*, *19*, 458–464.
- Belin, P., Zatorre, R. J., LaFalle, P., Ahead, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*, 309–312.
- Bohland, J. W., & Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. *Neuroimage*, *32*, 821–841.
- Brown, S., Ingham, R. J., Ingham, J. C., Laird, A. R., & Fox, P. T. (2005). Stuttered and fluent speech production: An ALE meta-analysis of functional neuroimaging studies. *Human Brain Mapping*, *25*, 105–117.
- 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., Martinez, M. J., & Parsons, L. M. (2006). Music and language side by side in the brain: A PET study of the generation of melodies and sentences. *European Journal of Neuroscience*, *23*, 2791–2803.
- Brown, S., Ngan, E., & Liotti, M. (2008). A larynx area in the human motor cortex. *Cerebral Cortex*, *18*, 837–845.
- Callan, D. E., Tsytarev, V., Hanakawa, T., Callan, A. M., Katsuhara, M., Fukuyama, H., et al. (2006). Song and speech: Brain regions involved with perception and covert production. *Neuroimage*, *31*, 1327–1342.
- Carson, R. G. (2005). Neural pathways mediating bilateral interactions between the upper limbs. *Brain Research Reviews*, *49*, 641–662.
- Catania, M., Jones, D. K., & Fitches, D. H. (2005). Perisylvian language networks of the human brain. *Annals of Neurology*, *57*, 8–16.
- Corfield, D. R., Murphy, K., Josephs, O., Fink, G. R., Frackowiak, R. S., Gu, A., et al. (1999). Cortical and subcortical control of tongue movement in humans: A functional neuroimaging study using fMRI. *Journal of Applied Physiology*, *86*, 1468–1477.
- Dromey, C., Ramig, L., & Johnson, A. B. (1995). Phonatory and articulatory changes associated with increased vocal intensity in Parkinson disease: A case study. *Journal of Speech and Hearing Research*, *38*, 751–764.
- Fónagy, I. (1981). Emotions, voice and music. *Research Aspects on Singing*, *33*, 51–79.
- Fónagy, I., & Magdics, K. (1963). Emotional patterns in intonation and music. *Zeitschrift für Phonetik*, *16*, 293–326.
- Fontaine, D., Capelle, L., & Duffau, H. (2002). Somatotopy of the supplementary motor area: Evidence from correlation of the extent of surgical resection with the clinical patterns of deficit. *Neurosurgery*, *50*, 297–305.
- Fox, P. T., Huang, A., Parsons, L. M., Xing, J.-H., Zamarippa, F., Rainey, L., et al. (2001). Location-probability profiles for the mouth region of the human primary motor-sensory cortex: Model and validation. *Neuroimage*, *13*, 196–209.
- Fox, P. T., & Lancaster, J. L. (2002). Mapping context and content: The BrainMap model. *Nature Reviews Neuroscience*, *3*, 319–321.
- Friston, K. J., Ashburner, J., Frith, C. D., Pauline, J.-B., Heather, J. D., & Frackowiak, R. S. J. (1995a). Spatial registration and normalization of images. *Human Brain Mapping*, *3*, 165–189.
- Friston, K. J., Holmes, A. P., Worley, K. J., Pauline, J.-B., Frith, C. D., & Frackowiak, R. S. J. (1995b). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, *2*, 189–210.
- Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, *15*, 870–878.
- Gerardin, E., Lehericy, S., Potion, J. B., Teasdale, S., Margain, J. F., Pompon, F., et al. (2003). Foot, hand, face and eye representation in the human striatum. *Cerebral Cortex*, *13*, 162–169.
- Grodd, W., Hülsmann, E., Lotze, M., Wildgruber, D., & Erb, M. (2001). Sensorimotor mapping of the human cerebellum: fMRI evidence of somatotopic organization. *Human Brain Mapping*, *13*, 55–73.
- Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: A PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex*, *9*, 697–704.
- Indefrey, P., & Level, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*, 101–144.
- Jeffries, K. J., Fritz, J. B., & Braun, A. R. (2003). Words in melody: An H₂¹⁵O PET study of brain activation during singing and speaking. *Neuroreport*, *14*, 749–754.
- Jordan, L. C., & Hillis, A. E. (2006). Disorders of speech and language: Aphasia, apraxia and dysarthria. *Current Opinion in Neurology*, *19*, 580–585.
- Jürgens, U. (2002). Neural pathways underlying vocal control. *Neuroscience and Biobehavioral Reviews*, *26*, 235–258.
- Kirzinger, A., & Jürgens, U. (1982). Cortical lesion effects and vocalization in the squirrel monkey. *Brain Research*, *233*, 299–315.
- Kleber, B., Birbaumer, N., Veit, R., Trevorrow, T., & Lotze, M. (2007). Overt and imagined singing of an Italian aria. *Neuroimage*, *36*, 889–900.
- Kochunov, P., Lancaster, J., Thompson, P., Toga, A. W., Brewer, P., Hardies, J., et al. (2002). An optimized individual target brain in the Talairach coordinate system. *Neuroimage*, *17*, 922–927.
- Ladd, D. R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.
- Laird, A. R., Fox, P. M., Price, C. J., Glahn, D. C., Uecker, A. M., Lancaster, J. L., et al. (2005b). ALE meta-analysis: Controlling the false discovery rate and performing statistical contrasts. *Human Brain Mapping*, *25*, 155–164.
- Laird, A. R., McMillan, K. M., Lancaster, J. L., Kochunov, P., Turkeltaub, P. E., Pardo, J. V., et al. (2005a). A comparison of label-based meta-analysis and activation likelihood estimation in the stroop task. *Human Brain Mapping*, *25*, 6–21.
- Lancaster, J. L., Tordesillas-Gutiérrez, D., Martínez, M., Salinas, F., Evans, A., Zilles, K., et al. (2007). Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. *Human Brain Mapping*, *28*, 1194–1205.
- Loucks, T. M., Poletto, C. J., Simonyan, K., Reynolds, C. L., & Ludlow, C. L. (2007). Human brain activation during phonation and exhalation: Common voluntary control for two upper airway functions. *Neuroimage*, *36*, 131–143.
- Maldjian, J. A., Laurienti, P. J., Kraft, R. A., & Burdette, J. H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage*, *19*, 1233–1239.
- Martin, R. E., MacIntosh, B. J., Smith, R. C., Barr, A. M., Stevens, T. K., Gati, J. S., et al. (2004). Cerebral areas processing swallowing and tongue movement are overlapping but distinct: A functional magnetic resonance imaging study. *Journal of Neurophysiology*, *92*, 2428–2443.
- Merati, A., Heman-Ackah, Y. D., Abaza, M., Altman, K. W., Sulica, L., & Belamowicz, S. (2005). Common movement disorders affecting the larynx: A report from the Neurology Committee of the AAO-HNS. *Otolaryngol. Head and Neck Surgery*, *133*, 654–665.
- Murphy, K., Corfield, D. R., Gu, A., Fink, G. R., Wise, R. J. S., Harrison, J., et al. (1997). Cerebral areas associated with motor control of speech in humans. *Journal of Applied Physiology*, *83*, 1438–1447.
- Ogar, J., Slama, H., Dronkers, N., Amici, S., & Gorno-Tempini, M. R. (2005). Apraxia of speech: An overview. *Neurocase*, *11*, 427–432.
- Özdemir, E., Norton, A., & Schlaug, G. (2006). Shared and distinct neural correlates of singing and speaking. *Neuroimage*, *33*, 628–635.
- Perry, D. W., Zatorre, R. J., Petrides, M., Alivisatos, B., Meyer, E., & Evans, A. C. (1999). Localization of cerebral activity during simple singing. *Neuroreport*, *10*, 3979–3984 (Also *Neuroreport*, *10*, 3452–3458).
- Petacchi, A., Laird, A. R., Fox, P. T., & Bower, J. M. (2005). Cerebellum and auditory function: An ALE meta-analysis of functional neuroimaging studies. *Human Brain Mapping*, *25*, 118–128.
- Ramig, L., Countryman, S., Thompson, L., & Horii, Y. (1995). A comparison of two forms of intensive speech treatment for Parkinson disease. *Journal of Speech and Hearing Research*, *38*, 1232–1251.
- Ramig, L., Sapir, S., Countryman, S., Pawlas, A., O'Brien, C., Hoehn, M., et al. (2001). Intensive voice treatment (LSVT) for individuals with Parkinson disease: A two-year follow-up. *Journal of Neurology, Neurosurgery and Psychiatry*, *71*, 493–498.
- Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*, 265–292.
- Riecker, A., Ackermann, H., Wildgruber, D., Dogil, G., & Grodd, W. (2000a). Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. *Neuroreport*, *11*, 1997–2000.
- Riecker, A., Ackermann, H., Wildgruber, D., Meyer, J., Dogil, G., Haider, H., et al. (2000b). Articulatory/phonetic sequencing at the level of the anterior perisylvian cortex: A functional magnetic resonance fMRI study. *Brain and Language*, *75*, 259–276.
- Riecker, A., Kassubek, J., Groschel, K., Grodd, W., & Ackermann, H. (2006). The cerebral control of speech tempo: Opposite relationship between speaking rate and BOLD signals changes at striatal and cerebellar structures. *Neuroimage*, *29*, 46–53.
- Riecker, A., Mathiak, K., Wildgruber, D., Erb, M., Hertrich, I., Grodd, W., et al. (2005). fMRI reveals two distinct cerebral networks subserving speech motor control. *Neurology*, *64*, 700–706.
- Riecker, A., Wildgruber, D., Dogil, G., Grodd, W., & Ackermann, H. (2002). Hemispheric lateralization effects of rhythm implementation during syllable repetitions: An fMRI study. *Neuroimage*, *16*, 169–176.
- Rotte, M., Kanowski, M., & Heinze, H. J. (2002). Functional magnetic resonance imaging for the evaluation of the motor system: Primary and secondary brain areas in different motor tasks. *Stereotactic and Functional Neurosurgery*, *78*, 3–16.
- Saito, Y., Ishii, J., Yagi, K., Tatsumi, I. F., & Mizusawa, H. (2006). Cerebral networks for spontaneous and synchronized singing and speaking. *Neuroreport*, *17*, 1893–1897.
- Sapir, S., Spielman, J., Ramig, L., Story, B., & Fox, C. (2007). Effects of intensive voice treatment LSVT on vowel articulation in dysarthric individuals with idiopathic Parkinson disease: Acoustic and perceptual findings. *Journal of Speech Language and Hearing Research*, *50*, 899–912.

- Schmahmann, J. D., Doyon, J., Toga, A. W., Petrides, M., & Evans, A. C. (2000). *MRI atlas of the human cerebellum*. San Diego: Academic Press.
- Simonyan, K., & Jürgens, U. (2002). Cortico-cortical projections of the motorcortical larynx area in the rhesus monkey. *Brain Research*, 949, 23–31.
- Simonyan, K., & Jürgens, U. (2003). Efferent subcortical projections of the laryngeal motorcortex in the rhesus monkey. *Brain Research*, 974, 43–59.
- Simonyan, K., & Jürgens, U. (2005a). Afferent subcortical connections into the motor cortical larynx area in the rhesus monkey. *Neuroscience*, 130, 119–131.
- Simonyan, K., & Jürgens, U. (2005b). Afferent cortical connections of the motor cortical larynx area in the rhesus monkey. *Neuroscience*, 130, 133–149.
- Simonyan, K., Saad, Z. S., Loucks, T. M., Poletto, C. J., & Ludlow, C. L. (2007). Functional neuroanatomy of human voluntary cough and sniff production. *Neuroimage*, 37, 401–409.
- Sörös, P., Guttman Sokoloff, L., Bose, A., McIntosh, A. R., Graham, S. J., & Stuss, D. T. (2006). Clustered functional MRI of overt speech production. *Neuroimage*, 32, 376–387.
- Sundberg, J. (1987). *The science of the singing voice*. Dekalb, IL: Northern Illinois University Press.
- Sutton, D., Larson, C., & Lindeman, R. C. (1974). Neocortical and limbic lesion effects on primate phonation. *Brain Research*, 71, 61–75.
- Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. New York: Thieme Medical Publishers.
- Terumitsu, M., Fujii, Y., Suzuki, K., Kwee, I. L., & Nakada, T. (2006). Human primary motor cortex shows hemispheric specialization for speech. *Neuroreport*, 17, 1091–1095.
- Turkeltaub, P. E., Eden, G. F., Jones, K. M., & Zeffiro, T. A. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: Method and validation. *Neuroimage*, 16, 765–780.
- Watanabe, J., Motoaki, S., Miura, N., Watanabe, Y., Yasuhiro, M., Matsue, Y., et al. (2004). The human parietal cortex is involved in spatial processing of tongue movement: An fMRI study. *Neuroimage*, 21, 1289–1299.
- Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, 7, 701–702.
- Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.