Experience-dependent neural substrates involved in vocal pitch regulation during singing

Jean Mary Zarate* and Robert J. Zatorre

Montréal Neurological Institute, McGill University, 3801 University Street, Montréal, Québec, Canada
International Laboratory for Brain, Music and Sound Research (BRAMS), Montréal, Québec, Canada

Received 16 August 2007; revised 21 December 2007; accepted 12 January 2008

Proper singing requires the integration of auditory feedback mechanisms with the vocal motor system, such that vocal pitch can be precisely controlled. To determine the neural substrates involved in audio-vocal integration, non-musicians and experienced singers underwent fMRI scanning while they sang a single tone with either unaltered (“simple”) or pitch-shifted auditory feedback; in pitch-shifted trials, subjects were instructed either to ignore or compensate for the shifted feedback. We hypothesized that the anterior cingulate cortex (ACC), superior temporal gyrus (STG), and anterior insula may be involved in audio-vocal integration due to their functional roles during singing and their anatomical connectivity. Although singers were more accurate than non-musicians in simple singing, both groups recruited similar functional networks. Singers ignored the shifted feedback better than non-musicians, and both groups also displayed different patterns of neural activity for this task: singers recruited bilateral auditory areas and left putamen, while non-musicians recruited the left supramarginal gyrus and primary motor cortex. While there were no significant group differences in performing the compensate task, singers displayed enhanced activity in the ACC, superior temporal sulcus, and putamen, whereas non-musicians exhibited increased activity in the dorsal premotor cortex, a region involved with sensorimotor interactions. We propose two neural substrates for audio-vocal integration: the dorsal premotor cortex may act as a basic interface, but with vocal training and practice, the ACC, auditory cortices, and putamen may be increasingly recruited as people learn to monitor their auditory feedback and adjust their vocal output accordingly.

© 2008 Elsevier Inc. All rights reserved.

Keywords: Audio-vocal integration; Auditory feedback; fMRI; Musicians; Pitch shift; Singers

Introduction

The control of vocal pitch plays an important role in both speech and song. Vocal pitch regulation requires the integration of both a stable vocal motor system and the auditory feedback system used to monitor vocal output. Although numerous studies have mapped out the neural mechanisms of the vocal motor system and investigated the auditory processing of voice, the components of a neural interface between these two systems are currently not well understood. Here, we present a functional magnetic resonance imaging (fMRI) study specifically designed to determine how auditory regions are integrated with the vocal motor system, with special emphasis on the control of vocal pitch. To gain more precise control over vocal pitch, experienced singers undergo many years of rigorous training. Therefore, we tested both non-musicians and experienced singers to test the hypothesis that improved vocal control in singers is accompanied by modulations within the network of regions recruited during singing.

Compared to speech, vocal pitch is more important in accomplishing the goals for singing, a task that constantly relies on auditory feedback and proper vocal control to produce specific, stable pitches; therefore, we used singing as a means to study audio-vocal integration. To date, there are only a few neuroimaging studies that focus on vocalization, and just a small portion of these have specifically investigated the neural substrates of singing. One of the earliest positron emission tomography (PET) studies of singing reported that the face area of the primary motor cortex, supplementary motor area, anterior cingulate cortex (ACC), insula, cerebellum, superior temporal gyrus (STG), and inferior parietal lobule were associated with repeatedly singing one note (Perry et al., 1999). In more recent singing studies, a similar functional network was recruited while subjects sang single notes, melodies, harmonized with a presented melody (Brown et al., 2004), or sang excerpts of an Italian aria (Kleber et al., 2007), which suggests that these regions may be part of a general functional network for singing. While earlier studies have investigated how singing and speech tasks are lateralized to different hemispheres of this functional network (Jeffries et al., 2003; Riecker et al., 2000;
neural correlates of speaking and singing within the singing
network, which include inferior sensorimotor cortex and auditory
areas (STG and superior temporal sulcus; Ozdemir et al., 2006). A
recent magnetoencephalography (MEG) study of vocalization also
found activity within sensorimotor cortices during singing and
speaking aloud (Gunji et al., 2007). Still, none of the previous
studies specifically investigated audio-vocal integration, which
would require manipulation of feedback. Two studies, however,
have suggested certain neural substrates for audio-vocal integration.
Hickok and colleagues (2003) used covert rehearsal of speech or
covert humming tasks and suggested that area “Spt” at the parietal–
temporal boundary plays a key role in audio-vocal integration, while
Schulz et al. (2005) used voiced and whispered speech tasks to
propose that the periaqueductal grey, anterior middle temporal
gyrus, and cerebellum are involved in integrating auditory
feedback with vocal control; yet, neither of these studies used
singing tasks, where audio-vocal integration is necessary to
monitor vocal pitch.

Audio-vocal integration requires communication between audi-
tory and vocal motor regions to ensure that vocalization matches its
intended output. Among the brain regions associated with simple
singing (Perry et al., 1999), the ACC, STG, and insula may be
candidates for audio-vocal integration, considering both their
functional properties and their connections with other brain regions.
The ACC has been associated with the initiation of vocalization in
primates (Jurgens, 1983; Muller-Preuss et al., 1980) and voluntary
speech and song in humans (Paus et al., 1993; Paus, 2001; Perry et al.,
1999; Picard and Strick, 1996), while the STG contains auditory
cortices that respond to various auditory stimuli, including speech
(Scott and Johnsrude, 2003). Electrophysiological and tracer studies
in primates have shown that during vocalization, the ACC inhibits
activity in auditory cortices via the extreme capsule (Barbas et al.,
1999; Muller-Preuss et al., 1980; Muller-Preuss and Ploog, 1981).
Prior to vocalization, these inhibitory influences suppress the majority
of neuronal activity in the STG to increase the cortical response to
incoming feedback (Eliades and Wang, 2003). The ACC also receives
input from the auditory cortices (Barbas et al., 1999; Jurgens, 1983;
Paus, 2001), which suggests that the ACC and STG may work
together to integrate auditory processing with vocal motor control.
The insula, which shares connections with the ACC and auditory
cortices (Mesulam and Mufson, 1982; Mufson and Mesulam, 1982),
may also provide an additional substrate for audio-vocal integration.
The majority of auditory cortices share numerous projections with the
mid-dorsal and posterior insula (Augustine, 1996), suggesting that
this region may be involved in auditory processing (Bamiou et al.,
2003; Mesulam and Mufson, 1982). However, the anterior insula, due
to its connections and cortical architecture (Rivier and Clarke, 1997),
possesses characteristics of a higher-order association area and may be
involved in cross-modal integration (Ackermann and Riecker, 2004;
Bamiou et al., 2003; Bushara et al., 2001; Lewis et al., 2000). In
accordance with this, the anterior insula was recruited during singing
and speaking aloud when compared to internal or covert vocalization
(Riecker et al., 2000); perhaps due to its connections with the ACC
and auditory cortices, the anterior insula may modulate ongoing vocal
processes by integrating auditory feedback with vocal output
(Ackermann and Riecker, 2004).

In the present experiment, subjects were asked to sing single
notes, and for certain tasks, the auditory feedback was pitch shifted
as they sang. We manipulated the feedback to identify the neural
substrates that integrate auditory feedback with vocal control,
using tasks modeled after those used by Burnett and colleagues
(Burnett et al., 1998; Burnett and Larson, 2002).

In our experiment, we presented two pitch-shift conditions that
required subjects to either ignore or compensate for the pitch shift;
this was done to mimic certain musical situations, as well as probe
the degree of cognitive control over the incoming feedback signal.
The ignore condition might be likened to the situation in choral
singing, where each singer must hold a different pitch. Singers must
ignore the other notes that are sung simultaneously and concentrate
on their intended note. For this condition, we hypothesized that
relative to normal singing, neural activity will increase in regions
associated with attentional control and decrease in auditory cortical
regions. The compensate condition simulated the situation in which
auditory feedback informs the singer that the produced pitch is sharp
or flat. As a result, the singer will voluntarily adjust the vocal output
until the auditory feedback matches the intended note. In this task, as
subjects attempt to fully correct for the pitch-shifted feedback, we
predict an increase in neural activity in the ACC, STG, and anterior
insula since subjects need to integrate auditory feedback and vocal
motor control to perform this task properly.

With practice, vocal output can be regulated more effectively—the
corrections of slight deviations should become more automatic. This
improvement in vocal control must be accompanied by changes in
neural processes. Previous research suggests that musical task-training
can induce experience-dependent changes in cortical activity, for
example, in tasks where co-activation of auditory and sensorimotor
areas is observed after learning (Bangert and Allenmuller, 2003; Lahav
et al., 2007). Investigations of experience-dependent cortical plasticity
may be especially facilitated by using musicians, who undergo years of
auditory and fine motor training to perfect their craft. For instance,
auditory cortices respond differentiably after extensive training (Pantev
et al., 2001a,b; Schneider et al., 2005). Similarly, anatomical changes
have been observed following musical training—larger volumes of the
anterior corpus callosum, right motor cortex, and cerebellum have
been reported in musicians (Schlaug, 2001).

While musician studies can elucidate neural mechanisms under-
lying cortical plasticity, trained or experienced singers, a subset of
musicians, have yet to be thoroughly investigated in such studies.
Trained singers may benefit from experience-dependent neural
plasticity since an early review suggests that extensive vocal
instruction and practice allow trained singers to sing accurately, even
when suffering from hearing loss (Wyke, 1974). Therefore, trained
singers may have better voice control in the absence of proper
feedback than non-musicians, and thus we predicted that experi-
enced singers would perform all three singing tasks more accurately
than non-musicians. Varying degrees of vocal pitch regulation may
be accompanied by different patterns of functional adaptations in the
brain. We hypothesized that experienced singers who constantly
monitor auditory feedback to regulate their voice would show an
enhancement or modulation in regions specifically involved in
audio-vocal integration, including ACC, STG, and anterior insula,
when compared to non-musicians.

Materials and methods

Subjects

A total of 31 healthy subjects were recruited from the McGill
University community. All subjects (mean age = 23 ± 5.09 years
old) were right-handed, had normal hearing, and were devoid of
neurological or psychological disorders and contraindications for
fMRI techniques. All subjects gave informed consent to participate in this study, in accordance with procedures approved by the Research Ethics Committees of the McConnell Brain Imaging Centre and the Montréal Neurological Institute.

Seven subjects were withdrawn from the study due to problems performing the vocal tasks, scheduling conflicts, or equipment failure. The remaining 24 subjects were categorized into 12 non-musicians and 12 experienced singers, according to self-report of vocal or musical experience. The six male and six female non-musicians had minimal musical exposure (an average of two years of training or experience) and did not currently play an instrument. The six male and six female experienced singers had an average of 10 years (±3.39 years) of formal vocal training or experience and were currently practicing or performing at the time of the study.

Experimental paradigm

All subjects underwent a familiarization session at least 1 day before testing in the scanner. For this session, each subject sat in front of a computer screen and a microphone (Audio-Technica ATM41HE, Stow, Ohio, USA) and wore headphones (Sony MDR-V900, New York, NY, USA), through which all auditory stimuli were delivered. The microphone was connected to a mixer to amplify the voice signal before it was sent to a VoiceOne pitch shifter (TC Helicon Vocal Technologies, Westlake Village, Calif., USA); during certain trials, the pitch shifter was programmed to alter the pitch of the vocal signal before it was delivered as auditory feedback to the subject's headphones. Throughout the entire session, pink noise was delivered through the headphones to reduce bone conduction, so that the manipulated vocal signal from the pitch shifter would be the main source of auditory feedback to the subjects. All auditory stimuli (pink noise, target vocal waves, and auditory feedback) were delivered to the headphones via the mixer (Fig. 1), and all volume levels were adjusted to comfortable levels for each subject. Pink noise was delivered at an average of 78.3 dB SPL A, while the target waves were presented at an average of 15.6 dB above the pink noise. The experiment was run using Media Control Functions (MCF) software (DigiVox, Montréal, Canada). Auditory feedback (via the pitch shifter) and all vocalizations were digitally recorded onto a Marantz PMD-670 digital recorder (Marantz Professional, Itasca, Illinois, USA) for subsequent analysis.

In the familiarization session, subjects practiced three different conditions, all of which were presented again during the fMRI scanning session. For all singing tasks, we first presented a target note and then used a visual cue to prompt the subjects to sing the note back. All subjects were trained to sing using the syllable “ah” with minimal mouth movement to reduce movement artifacts in the fMRI session. They were instructed to keep their jaws slightly open and lips closed, so that at the beginning and end of every sung note, only their lips, but not their jaws, moved. Each singing condition was presented in blocks of five trials, with a different note for each trial so that we spanned a range of pitches (D#3, F3, G#3, B3, C#4 for males; an octave higher for females). In the first task, after hearing the 2-s target note, subjects were cued to sing the note for 4 s (“simple” condition). In the other tasks, approximately 1 s after the onset of singing (shift onset range, 1000–1500 ms), the voice was shifted either up or down by 200 cents (one whole tone) via the pitch shifter and remained shifted until the end of the trial (“pitch-shift” conditions); this large shift interval was chosen to ensure that all non-musicians would detect a change in auditory feedback. For these pitch-shift conditions, subjects were instructed to make a different response in each of two distinct conditions: (1) ignore the shifted feedback and keep the vocal output as steady as possible on the original note (“ignore”, Fig. 2a) or (2) correct for the shifted feedback so that the feedback sounded like the target note (“compensate”, Fig. 2b). We maintained the feedback shift until the end of the trial to increase the probability of finding brain regions involved in vocal pitch regulation with fMRI techniques. During familiarization, the subjects went through four experimental runs with all conditions included in each run.

At least one day after the familiarization session, each subject was tested in a Siemens Sonata 1.5-T magnetic resonance (MR) scanner. Each subject was fitted with MR-compatible headphones with an attached MR-compatible microphone (Commander XG headset, Resonance Technology, Inc., Northridge, Calif., USA). The audio setup for fMRI testing was the same as the behavioral setup described above. Sound levels were again adjusted to comfortable levels for each subject before the experiment began. Pink noise was presented at an average of 69.1 dB SPL A, and the target notes were

Please cite this article as: Zarate, J.M., Zatorre, R.J., Experience-dependent neural substrates involved in vocal pitch regulation during singing. NeuroImage (2008), doi:10.1016/j.neuroimage.2008.01.026
presented at an average of 10.4 dB above the pink noise. All visual cues were back-projected onto a screen at the subjects’ feet, viewed via a mirror attached to the head coil. While in the scanner, subjects were exposed to the same three singing conditions presented in the familiarization session. Two control conditions were also presented: (1) a condition with only pink noise playing in the background, used as a baseline, and (2) a perception task, which presented a target note that subjects did not have to sing back, thus serving as an auditory control for all singing conditions in the scanner. In both of these control conditions, subjects were visually cued to breathe out normally rather than sing; therefore, these conditions also served as a respiratory control for the singing conditions. Prior to functional scanning, a high-resolution (voxel size = 1 mm$^3$) T1-weighted scan was obtained for anatomical localization. During the two functional runs, one whole-head frame of 25 contiguous T2*-weighted images were obtained for anatomical localization. During the two functional runs, one whole-head frame of 25 contiguous T2*-weighted images were acquired in an interleaved fashion (TE = 85 ms, TR = 10 s, 64 × 64 matrix, voxel size = 5 mm$^3$, FOV = 320 mm$^2$). We utilized a sparse-sampling, event-related experimental design, in which tasks were performed during the silences between image acquisitions to prevent scanner noise from interfering with the auditory stimuli and to reduce any effect of movement (Belin et al., 1999). Timings of task presentations were systematically varied or “jittered” by ± 500 ms to maximize the likelihood of obtaining the peak of the hemodynamic response for each task. Within each run, each of the singing and control conditions was presented 10 times, with one scan acquisition per condition. Each subject went through two experimental runs in the scanner, resulting in a total of 20 acquisitions per condition. The order of the conditions within each run was counterbalanced across subjects, and the five target note presentations were pseudo-randomized within each condition.

**Behavioral analyses**

Prior to frequency extraction, the vocalizations recorded in the MR scanner were notch-filtered to better isolate fundamental frequencies ($f_0$) and remove all extraneous frequencies due to environmental noise (males: 100 Hz > $f_0$ > 500 Hz filtered out; females: 200 Hz > $f_0$ > 750 Hz filtered out). The familiarization session recordings were not filtered since they were taken in a quiet, sound-attenuated booth. The $f_0$ was extracted from all vocal recordings using Praat, a speech-analysis program (Boersma and Weenink, 2004). All frequencies were converted from absolute frequency in Hertz to relative frequency from the target note in cents (100 cents = one semitone) with the following formula: $100 \times \log_{10}(f/f_0)$, where $f_0$ equals the frequency of the produced note, and $f_2$ is the frequency of the target note. This conversion normalized the vocal data for comparison across genders, individual pitches, and subject groups.

After initial processing, the resulting vocal data from the three singing tasks were handled differently. To reduce variability both within and across simple singing trials from different vocalization onset and offset times, the first 500 ms and last 50–650 ms were excluded from the data; the remaining vocal data corresponded to a vocalization length of 2850–3000 ms, which was typically the most stable portion of the sung output. All data from pitch-shifted singing tasks were aligned at the onset of the pitch shift and divided into either pre-shift or post-shift data. To reduce variability due to onset and offset times, pre-shift singing data were truncated to 150–500 ms immediately before the pitch shift, and post-shift singing data were truncated to 1900–3000 ms after the pitch shift. Since the target response magnitude for the ignore task is 0 cents (i.e., no vocal response to the pitch shift), no further calculations were performed on the vocal data for this task. We had no a priori hypotheses about differences in brain activity while ignoring or compensating for pitch shifts in different directions; therefore, all pitch-shifted data were transformed into a uniform template that entailed ignoring or compensating for a downward shift of 200 cents (i.e., ~200 cents). Within each compensate trial, the magnitude of the pitch-shift response was calculated relative to the average $f_0$ during pre-shift singing (see Fig. 2b).

Each singing condition was analyzed separately in two-way repeated-measures analyses of variance (ANOVAs, group by pitch). To determine the effects of vocal training or different target pitches on vocal accuracy, we used the average error from the target note (in cents) as the dependent variable for the simple singing task, and we used the average magnitude of the response to the pitch shift (in
cents) as the dependent variable for pitch-shifted tasks. Simple effects tests were used to analyze significant interactions, and the Tukey's honest significant difference test was used for post hoc analyses. We also analyzed the standard deviations of the subjects' performances during all singing tasks to see if subjects' vocal stability varied as feedback was altered (i.e., smaller standard deviations reflect greater vocal stability).

**fMRI analyses**

To correct for motion artifacts, all blood-oxygen-level-dependent (BOLD) images within each run were realigned with the third frame using the AFNI software (Cox, 1996). To increase the signal-to-noise ratio of the imaging data, the images were then spatially smoothed (or low-pass filtered) with an 8-mm full-width at half-maximum (fwhm) isotropic Gaussian kernel. Prior to analysis, frames one through three and frame 39 were excluded from further analyses because they were either practice singing trials or frames in which the “ignore” or “compensate” instructions were presented. Image analyses were performed with FMRISTAT, which involves a series of MATLAB scripts that utilize the general linear model for analyses (Worsley et al., 2002). The general linear model $Y = \beta X + \epsilon$ expresses the response variable (BOLD signal) $Y$ in terms of a linear combination of the explanatory variable (stimulus) $X$, the parameter estimates (effects of interest) $\beta$, and the error term $\epsilon$. Temporal drift is modeled as cubic splines and then removed since it can be confounded with the BOLD response. The first MATLAB script fmridesign created the design matrix within each run, where each column contained the explanatory variables and each row represented a scan. The program fmri2im then fitted the linear model with the fMRI time series and solved for the parameter estimates $\beta$ with a least-squares solution, yielding estimates of effects, standard errors, and $t$ statistics for each contrast and for each run. An effect of interest was specified by a vector of contrast weights that gave a weighted sum or compound of parameter estimates referred to as a contrast. We entered our planned comparisons into the analyses (e.g., each singing condition versus perception task, comparisons between the singing conditions). The third MATLAB program multistat combined runs together within subjects, using a fixed-effects general linear model. Before group statistical maps for each contrast of interest were generated, in-house software was used to linearly transform anatomical and functional images from each subject into standard MNI/Talairach stereotaxic coordinate space, using the MNI 305 template (Collins et al., 1994; Talairach and Tournoux, 1988).

Then, we specified a mixed-effects linear model in multistat when averaging data across subjects; the data were smoothed with an fwhm Gaussian filter so that the ratio of the random-effects variance divided by the fixed-effects variance results in approximately 100 degrees of freedom. Lastly, the program stat_summary reported the minimum $p$-values among those computed with a Bonferroni correction, random field theory, and discrete local maxima (Worsley, 2005). We reported peaks that surpassed an uncorrected $p$-value of 0.001 ($t = 3.17$). However, we also indicated which peaks survived the critical $t$-threshold corrected for multiple comparisons, which $stat.-summary$ determined to be 5.1 at $p = 0.05$, using a whole-brain search volume.

For functional connectivity, the general linear model was fitted to account for the neural activity due to a stimulus (e.g., any singing task) and a baseline (perception). Then, the remaining activity within a specific voxel (the “seed” voxel) was regressed on the activity within the rest of the brain (on a voxel-by-voxel basis) to determine where activity significantly covaries with the activity in that seed voxel, without the effect of a stimulus (Friston, 1994; Worsley et al., 2005). Due to the exclusion of many pitch-shifted trials (as a result of equipment failure or subject-performance error), we did not have enough data to properly perform functional connectivity analyses for the ignore and compensate tasks. Therefore, we reported functional connectivity results using the simple singing condition as the stimulus and the perception task as a baseline. For this analysis, all 24 subjects were combined to increase statistical power, since the differences in neural activity between experienced singers and non-musicians were very subtle for simple singing. We chose the left STG, right mid-dorsal insula, right STG, and left ACC as seed voxels because they were located in or near the hypothesized substrates for audio-vocal integration and were also significantly active following a subtraction analysis that compared simple singing with perception in all 24 subjects. The $t$-map from this subtraction analysis was used to construct a mask containing only the positive voxels with values less than or equal to an uncorrected $p$-value of 0.001. We applied this mask to functional connectivity contrasts to constrain the search for significant voxels to those only within this functional network, which should be the network recruited during all singing tasks.

Conjunction analyses were performed for the “simple singing versus perception”, “ignore versus simple singing”, and “compensate versus simple singing” contrasts. For each contrast, we found the minimum $t$-statistic at each voxel across the images for both groups. The conjunction results were then tested against the “conjunction null hypothesis”, which entailed using the critical threshold for just one contrast to determine whether there was significant neural activity in certain brain regions in either or both groups (Nichols et al., 2005).

Voxel-of-interest (VOI) analyses were performed on voxels that displayed peak activity in group-contrasted BOLD images. For each voxel in MNI space, the BOLD signal is extracted from the corresponding voxel in native (non-standardized) space within each subject. At each VOI, the BOLD signal for the task-of-interest is compared to the BOLD signal during the baseline condition in the following way: [(BOLD signal at task−BOLD signal at baseline)/BOLD signal at baseline]).

**Data exclusions**

For behavioral analyses, 162 out of 1920 familiarization recordings and 109 out of 960 fMRI recordings were excluded from analyses due to equipment failure, subject-performance error, or problems with frequency extraction. For fMRI analyses, 72 out of 1920 frames were excluded from analyses due to equipment failure or performance errors only.

**Results**

**Behavioral results from familiarization session**

**Accuracy: average error from target notes or target response**

The ANOVA performed on the simple singing data revealed a significant pitch main effect $F(4,88)=5.15, p<0.001$ and a two-way interaction that showed a trend towards significance $F(4,88)=2.18, p<0.08$; the group main effect was not significant ($p>0.1$). Post hoc analyses of the pitch main effect determined that both groups sang much flatter on the highest two notes when compared to the lowest note (not shown, all $p<0.01$). Simple effects tests and post hoc
Fig. 3. Vocal accuracy of non-musicians (NON-MUS) and experienced singers (SINGER) during the familiarization and fMRI sessions. (a) Vocal accuracy, measured as error or response magnitude (+SE) in cents, across five target notes in all singing tasks during the familiarization session. Perfect vocal accuracy is depicted as 0 cents in both the simple and ignore singing conditions, while perfect vocal accuracy is 200 cents in the compensate task. During the simple singing task, non-musicians sang the two highest notes with less accuracy than the two lowest notes during (denoted by *, p < 0.05), and they also sang the highest note less accurately than experienced singers (denoted by !, p < 0.05). During the ignore task, non-musicians were also less accurate at ignoring shifted feedback across all notes, when compared with singers (denoted by !, p < 0.05). (b) Vocal accuracy across five target notes in all singing tasks during the fMRI session. Within the non-musicians, the highest three notes were less accurate than the lowest note (denoted by *, p < 0.05), and the highest two notes were less accurate than the second lowest note (denoted by !, p < 0.01) in the simple singing task. During the ignore task, non-musicians were also less accurate at ignoring shifted feedback across all notes, when compared with singers (denoted by !, p < 0.05).
analyses performed on the two-way interaction revealed that experienced singers were more accurate on the highest note than non-musicians \(p < 0.05\), and among the non-musicians, the highest two notes were sung less accurately than the lowest two notes \(p < 0.05\).

Since non-musicians exhibited such inaccurate performances in simple singing, we subsequently normalized all of the pitch-shifted vocal data to correct for subjects' vocal inaccuracy (i.e., absolute pitch level). After this correction, statistical analyses were performed on the magnitudes of vocal responses to pitch-shifted auditory feedback. During the ignore task, the ANOVA revealed a significant group main effect—the non-musicians were generally less accurate at ignoring shifted feedback than singers \(F(1,22) = 7.63, p < 0.05\). In the compensate task, no significant main effects or interactions were found, indicating that there were no significant differences in performance as both groups corrected for shifted feedback across all notes \(p < 0.05\).

**Stability: standard deviation across vocalizations**

The ANOVA on stability measures from the simple singing condition resulted in a group main effect \(F(1,22) = 8.57, p < 0.001\)—non-musicians displayed more vocal instability than singers. Similar to the accuracy results in the ignore task, a significant group main effect was revealed while subjects were ignoring shifted feedback; non-musicians were less stable across all notes than singers \(p = 0.001\). In the compensate task, no significant main effects or interactions were found \(p < 0.05\).

**Behavioral results from fMRI session**

**Accuracy: average error from target notes or target response**

Analyses of the simple singing data revealed a significant pitch main effect \(F(4,88) = 5.95, p < 0.001\) and a group-by-pitch interaction \(F(4,88) = 4.94, p < 0.01\); once again, the group main effect was not significant \(p > 0.6\). Simple effects tests and post hoc analyses of the interaction determined that non-musicians sang the highest three notes much flatter than the lowest two notes \(p < 0.05\), while there was no significant difference in the experienced singers' accuracy across all notes. For the ignore task, the ANOVA revealed a significant group main effect—across all pitches, non-musicians were less successful at ignoring shifted feedback than singers \(F(1,20) = 7.69, p < 0.05\). Similar to the results in the familiarization session, no significant main effects or interaction were seen in the compensate task \(p < 0.05\).

---

**Fig. 4.** Measures of vocal stability (average standard deviation \(+SE\), in cents) in non-musicians (NON-MUS) and experienced singers (SINGER) averaged across all target notes within each task. Larger numbers reflect greater vocal instability. (a) Vocal stability during the familiarization session. Non-musicians showed more vocal instability than experienced singers in both the simple and ignore singing tasks (denoted by \(\dagger\), \(p < 0.01\)). (b) Vocal stability during the fMRI session. Similar to the familiarization session results, non-musicians also exhibited more vocal instability in both the simple and ignore singing tasks than experienced singers (denoted by \(*\), \(p < 0.05\)).

Please cite this article as: Zarate, J.M., Zatorre, R.J., Experience-dependent neural substrates involved in vocal pitch regulation during singing. *NeuroImage* (2008), doi:10.1016/j.neuroimage.2008.01.026
Stability: standard deviation across vocalizations

The ANOVA on stability measures from the simple singing condition resulted in a significant group main effect, where non-musicians displayed less vocal stability across all notes than singers [Fig. 4b, left panel; \( F(1,22) = 5.47, p < 0.05 \)]. Non-musicians also showed less vocal stability than singers during the ignore task [Fig. 4b, middle panel; \( F(1,20) = 5.35, p < 0.05 \)]. In the compensate task, no significant main effects or interactions were seen (Fig. 4b, right panel).

fMRI results

In the following section, we will use the fMRI data from the non-musicians to outline the basic functional network associated with the simple singing task; this network should be common to most people in the population who do not undergo extensive vocal training or experience. We will then present how the ignore and compensate tasks affect this basic network. Finally, we will examine how extensive vocal experience affects this functional network in each of the singing tasks. A preliminary version of these data was previously reported in a short, poster-based publication (Zarate and Zatorre, 2005).

Basic functional network for singing in non-musicians

Simple singing, when compared with the perception task, recruited a number of regions in non-musicians, which included left anterior cingulate cortex (ACC), bilateral primary and secondary auditory areas, bilateral supplementary motor areas (SMA), bilateral primary motor cortices (M1), right mid-dorsal insula, bilateral somatosensory cortices, thalamus, and bilateral cerebellum (Fig. 5, Table 1).

Fig. 5. Functional networks for all singing tasks in non-musicians (see Tables 1–3 for specific coordinates and additional regions). **Left column**: Functional network of regions associated with singing single notes, when compared with perception. In coronal and horizontal images, the left side of the image is the left side of the brain. **Middle column**: The ignore task, when compared to simple singing, recruited right ACC, left intraparietal sulcus, left supramarginal gyrus, and right anterior insula. **Right column**: When compared to the simple singing network, the compensate task recruited the right ACC, left intraparietal sulcus, left supramarginal gyrus, and left dorsal premotor cortex. ACC, anterior cingulate cortex; Cbl, cerebellum; dPMC, dorsal premotor cortex; INS, insula; IPS, intraparietal sulcus; M1, primary motor cortex; PAC, primary auditory cortex; SMA, supplementary motor area; SMG, supramarginal gyrus; STG, superior temporal gyrus; Th, thalamus.

Please cite this article as: Zarate, J.M., Zatorre, R.J., Experience-dependent neural substrates involved in vocal pitch regulation during singing, NeuroImage (2008), doi:10.1016/j.neuroimage.2008.01.026
Additional brain regions recruited for singing with pitch-shifted auditory feedback

When the ignore task was contrasted with the simple singing condition in non-musicians, we found that right rostral ACC, right dorsal premotor cortex, left dorso- and ventrolateral prefrontal cortex, left intraparietal sulcus, left supramarginal gyrus, and right anterior insula were recruited with the basic singing network to ignore pitch-shifted feedback while singing (Fig. 5, Table 2). The contrast between the compensate and simple singing tasks in non-musicians revealed that correcting for shifted auditory feedback recruited left posterior STG, right ACC, left dorsal premotor cortex, left dorsolateral prefrontal cortex, left intraparietal sulcus, and left supramarginal gyrus, in addition to the network of regions used for simple singing (Fig. 5, Table 3).

Similarities and experience-dependent changes in the basic functional network for singing

The contrasts described above (simple singing versus perception, ignore versus simple singing, and compensate versus simple singing) were also performed with images collected from the 12 experienced singers (see Tables 1–3, Fig. 6). The data from both groups were entered into between-group conjunction analyses to find group similarities in the patterns of neural activity during each condition. Additionally, the group images were statistically compared to determine group differences.

The conjunction analysis performed on the “simple singing versus perception” contrast confirmed that both non-musicians and experienced singers recruited the basic singing network during simple singing, i.e., bilateral auditory cortices, left ACC, bilateral SMA and M1 mouth regions, right ventral premotor cortex, right mid-dorsal insula, left somatosensory cortex, thalamus, and cerebellum (Table 1). In the between-group comparisons, singers displayed more neural activity in the left angular gyrus and left frontopolar gyrus (not shown). On the other hand, non-musicians exhibited more neural activity in the left cuneus (not shown). However, VOI analyses at each of these regions revealed that the group differences were mainly attributed to a larger decrease in BOLD signal in one group or the other when compared to baseline neural activity [decrease seen in the angular and frontopolar gyri for non-musicians; seen in cuneus for singers (not shown)]. Therefore, we conclude that a similar basic functional singing network is recruited in both groups, regardless of singing experience.

In the “ignore versus simple singing” contrast, the conjunction analysis showed that both groups displayed increased activity within right dorsal premotor cortex, right anterior insula, left intraparietal sulcus, and bilateral supramarginal gyrus as subjects ignored the

Table 1: Functional network associated with simple singing

<table>
<thead>
<tr>
<th></th>
<th>Non-musicians</th>
<th>Singers</th>
<th>Conjunction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Left (x y z t)</td>
<td>Left (x y z t)</td>
<td>Left (x y z t)</td>
</tr>
<tr>
<td><strong>Auditory</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAC</td>
<td>–54 –18 6 8 7.3</td>
<td>–40 –24 0 4.4</td>
<td>–42 –28 6 4.6</td>
</tr>
<tr>
<td>STG</td>
<td>–54 8 0 5.3 60</td>
<td>–40 16 12 6.5</td>
<td>–42 –34 12 5.9</td>
</tr>
<tr>
<td>Planum temporale</td>
<td>–38 –34 16 6.0</td>
<td>–34 18 5.3</td>
<td>–34 16 4.2</td>
</tr>
<tr>
<td><strong>Motor</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACC (BA 24)</td>
<td>–2 4 42 5.4</td>
<td>–2 0 46 3.9</td>
<td></td>
</tr>
<tr>
<td>SMA</td>
<td>0 –4 58 4.2</td>
<td>2 –58 4.5</td>
<td></td>
</tr>
<tr>
<td>M1 (mouth region)</td>
<td>–46 14 44 4.7</td>
<td>–46 14 44 4.7</td>
<td></td>
</tr>
<tr>
<td>vPMC</td>
<td>48 4 26 4.1</td>
<td>–58 4 4 5.1</td>
<td>62 6 28 5.6</td>
</tr>
<tr>
<td><strong>Multimodal</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insula</td>
<td>38 –16 –6 4.3</td>
<td>38 –2 0 4.2</td>
<td></td>
</tr>
<tr>
<td><strong>Parietal</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Somatosensory</td>
<td>–38 –20 40 4.9</td>
<td>–40 –20 40 4.9</td>
<td></td>
</tr>
<tr>
<td><strong>Subcortical</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thalamus</td>
<td>0 –22 14 4.9</td>
<td>6 –4 4 4.1</td>
<td>–2 –26 8 4.3</td>
</tr>
<tr>
<td>Lateral globus pallidus</td>
<td>22 –2 –6 5.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cerebellum</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VI (declive)</td>
<td>–24 –60 –22 5.1</td>
<td>–26 –64 –24 3.7</td>
<td></td>
</tr>
<tr>
<td>IV (culmen)</td>
<td>0 –48 –4 5.1</td>
<td>–26 –64 –24 3.7</td>
<td></td>
</tr>
<tr>
<td>V (culmen)</td>
<td>2 –54 –26 4.2</td>
<td>2 –56 –24 3.7</td>
<td></td>
</tr>
</tbody>
</table>

Regions of peak neural activity associated with simple singing (compared with perception). Coordinates are reported in MNI305/Talairach space. ACC, anterior cingulate cortex; BA, Brodmann area; M1, primary motor cortex; PAC, primary auditory cortex; SMA, supplementary motor area; STG, superior temporal gyrus; vPMC, ventral premotor cortex. r = 3.17, p = 0.001 (uncorrected). Peaks that survived the critical corrected r = 5.1, p < 0.05 are marked with an asterisk. Marginally significant peaks are marked with parentheses.

Please cite this article as: Zarate, J.M., Zatorre, R.J., Experience-dependent neural substrates involved in vocal pitch regulation during singing, NeuroImage (2008), doi:10.1016/j.neuroimage.2008.01.026
shifted feedback (Table 2). The between-group comparison determined that singers recruited more activity within the left putamen, bilateral STG, and right superior temporal sulcus (STS) than non-musicians during the ignore task (Fig. 7a, Table 4a). In contrast, non-musicians showed increased activity within the left supramarginal gyrus and the mouth region of M1 as they ignored the pitch shift (Table 4a).

For the “compensate versus simple singing” contrast, the conjunction analysis determined that both groups recruited the right rostral ACC, left dorsal premotor cortex, bilateral intraparietal sulcus, and left supramarginal gyrus when correcting for pitch-shifted feedback (Table 3). The between-group comparison revealed that singers showed stronger activity in left putamen, right STS, left rostral ACC, and right M1 than non-musicians while correcting for the pitch shift (Fig. 7b, Table 4b). On the other hand, non-musicians displayed stronger activity in subparietal sulcus, superior occipital cortex, and dorsal premotor cortex during the compensate task (Table 4b).

Functional connectivity within the simple singing network

Here, we only report the functional connectivity results for the simple singing task in both groups combined (see Materials and methods, fMRI analyses). We chose seed voxels within our hypothesized regions (i.e., left ACC, bilateral STG, and right insula; see Materials and methods, fMRI analyses for details) to determine if they form a functional network during each singing task. Fig. 8 displays the connectivity maps for the left STG and left ACC seed voxels, as well as the overlap in the ACC, bilateral STG, and right insula. This suggests that the hypothesized regions are functionally connected with each other, with the exception of the right STG, where we did not find any significant correlations between the right STG and ACC (Table 5). Furthermore, we found that most of the brain regions within the functional network for singing (i.e., additional auditory, motor, premotor areas and thalamus), as revealed by the subtraction analyses for simple singing, were also functionally correlated with each other (Table 5).

Discussion

Behavioral results

As expected, non-musicians were both less accurate and less vocally stable than singers in the simple singing task. Similarly, in the ignore task, non-musicians were also less accurate and less vocally stable than singers because they partially corrected for the pitch shift, rather than ignoring it completely. In this task, vocal inaccuracy may be due to a prepotent pitch-shift response. Previous behavioral studies have demonstrated that during vocalization, tiny pitch shifts in auditory feedback elicit compensatory pitch-shift responses, reflecting adjustment in the opposite direction of the pitch shift (Burnett et al., 1998; Burnett and Larson, 2002; Jones and Munhall, 2000, 2005). Furthermore, as a pitch shift increases in duration, two pitch-shift responses often emerge: a compensatory, early pitch-shift response that is quick and fairly automatic, and a late pitch-shift response that is under partial voluntary control (Burnett et al., 1998; Hain et al., 2000). After correcting for the non-musicians’ inaccurate initial reproduction of the target notes,

Table 2

<table>
<thead>
<tr>
<th></th>
<th>Non-musicians</th>
<th></th>
<th>Singers</th>
<th></th>
<th>Conjunction</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Left</td>
<td>Right</td>
<td>Left</td>
<td>Right</td>
<td>Left</td>
<td>Right</td>
</tr>
<tr>
<td></td>
<td>x y z t</td>
<td>x y z t</td>
<td>x y z t</td>
<td>x y z t</td>
<td>x y z t</td>
<td>x y z t</td>
</tr>
<tr>
<td><strong>Auditory</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAC</td>
<td></td>
<td></td>
<td>-52</td>
<td>-18</td>
<td>4</td>
<td>4.2</td>
</tr>
<tr>
<td>STG</td>
<td></td>
<td></td>
<td>-62</td>
<td>-30</td>
<td>10</td>
<td>5.0</td>
</tr>
<tr>
<td>Upper bank STS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>64</td>
<td>-34</td>
</tr>
<tr>
<td>Lower bank STS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td>-44</td>
</tr>
<tr>
<td>Planum temporale</td>
<td></td>
<td></td>
<td>-50</td>
<td>-32</td>
<td>6</td>
<td>4.5</td>
</tr>
<tr>
<td><strong>Motor</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACC (BA 32)</td>
<td>8</td>
<td>18</td>
<td>40</td>
<td>4.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dPMC</td>
<td>-4</td>
<td>52</td>
<td>3.8</td>
<td></td>
<td>-4</td>
<td>56</td>
</tr>
<tr>
<td>vPMC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-4</td>
<td>54</td>
</tr>
<tr>
<td><strong>Multimodal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insula</td>
<td>36</td>
<td>20</td>
<td>-2</td>
<td>4.0</td>
<td>40</td>
<td>20</td>
</tr>
<tr>
<td>VLPFC-BA 44</td>
<td>-54</td>
<td>12</td>
<td>28</td>
<td>3.9</td>
<td>36</td>
<td>20</td>
</tr>
<tr>
<td><strong>Parietal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intraparietal sulcus</td>
<td>-48</td>
<td>-50</td>
<td>56</td>
<td>4.7</td>
<td>-54</td>
<td>-36</td>
</tr>
<tr>
<td>Supramarginal gyrus</td>
<td>-52</td>
<td>-34</td>
<td>30</td>
<td>4.4</td>
<td>-46</td>
<td>-40</td>
</tr>
<tr>
<td>Angular gyrus</td>
<td>-38</td>
<td>-54</td>
<td>50</td>
<td>4.6</td>
<td>62</td>
<td>-30</td>
</tr>
<tr>
<td>IPL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>62</td>
<td>46</td>
</tr>
</tbody>
</table>

Regions of peak neural activity during the ignore singing task, when compared with simple singing. DLPFC, dorsolateral prefrontal cortex; dPMC, dorsal premotor cortex; IPL, inferior parietal lobule; STS, superior temporal sulcus; VLPFC, ventrolateral prefrontal cortex. $t=3.17, p=0.001$ (uncorrected). Peaks that survived the critical corrected $t=5.1, p<0.05$ are marked with an asterisk. Marginally significant peaks are marked with parentheses.
**Table 3**

| Additional brain regions associated with singing while compensating for pitch-shifted feedback |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|
|                                 | Non-musicians    | Singers         | Conjunction      |
|                                 | Left             | Right            | Left             | Right            |
|                                 | x    y   z  t    | x    y   z  t    | x    y   z  t    | x    y   z  t    |
| **Auditory**                    |                  |                  |                  |                  |
| STG                             | -62  -44  20  3.6 | -56  -32  4  3.9 | 48  -34  0  4.9  |
| Upper bank STS                  |                  |                  |                  |                  |
| Lower bank STS                  |                  |                  |                  |                  |
| **Motor**                       |                  |                  |                  |                  |
| ACC (BA 32)                     | 8    6    46  4.0 | -4   24   30  3.6 | 8    6    48  3.9 |
| pre-SMA                         |                  |                  |                  |                  |
| dPMC                            | -34  -14  70  3.9 | -44  -4   54  4.2 | 52  2    46  4.4  |
| vPMC                            |                  |                  |                  |                  |
| **Multimodal**                  |                  |                  |                  |                  |
| Insula                          |                  |                  |                  |                  |
| **Frontal**                     |                  |                  |                  |                  |
| DLPCF                           | -34  -2   64  3.6 |                  | 28   -6   52  4.3 |
| VLPFC                           |                  |                  |                  |                  |
| **Parietal**                    |                  |                  |                  |                  |
| Intraparietal sulcus            | -50  -48  58  5.0 |                  | 42  -50  60  4.0  |
| Supramarginal gyrus             | -54  -40  26  4.2 | -56  -42  50  4.2 | 58  -30  42  4.3  |
| **Subcortical**                 |                  |                  |                  |                  |
| Thalamus                        |                  |                  |                  |                  |
| Putamen                         | -28  2    6  3.9  |                  | 26   8    4  4.3  |
| Lateral globus pallidus         |                  |                  |                  |                  |

Regions of peak neural activity during the compensate singing task, when compared with simple singing. pre-SMA, pre-supplementary motor area. *t*=3.17, *p*<0.001 (uncorrected). Peaks that survived the critical corrected *t*=5.1, *p*<0.05 are marked with an asterisk.

We did not find any significant group differences in compensating for the pitch shift. This response may be a “default” reaction to pitch-shifted feedback, much like other compensatory motor responses to perturbations to vocal amplitude [i.e., the Lombard effect (Lombard, 1911; Siegel and Pick, 1974)], changed formant frequencies during speech production (Houde and Jordan, 1998; Houde and Jordan, 2002; Purcell and Munhall, 2006a,b), or muscular adjustments after an unexpected load perturbation (Winstein et al., 1991). This “default” compensatory response may fall under more complete voluntary control after training and practice, as seen in our experienced singers in the ignore task. Functional imaging results

**Network for simple singing**

Although they differed in both accuracy and vocal stability during the simple singing condition, both groups recruited similar functional networks, which included bilateral auditory areas, primary motor and premotor areas (ACC and SMA), right mid-dorsal insula, somatosensory areas, thalamus, and cerebellum. This network is similar to the functional network for singing reported in other studies (Kleber et al., 2007; Perry et al., 1999). Since a similar pattern of neural activity was seen in all of these singing studies, regardless of task complexity, we consider the network associated with simple singing as the basic functional network recruited during all singing tasks in this experiment.

While subtraction analyses may show which brain regions are active during singing tasks, we cannot conclude from this alone that these regions form a singing-related network. However, functional connectivity analyses determined that neural activity is indeed correlated among most of the regions found within the functional network (i.e., bilateral auditory areas, ACC, insula, motor and premotor areas, and thalamus), which has not been established in previous neuroimaging studies of singing. While these analyses do not show how these regions may be causally linked, they provide evidence that these regions are not only recruited during singing, but also that they specifically interact with each other to perform the singing tasks.

Audio-vocal integration in pitch-shifted tasks

In the ignore and compensate tasks, both groups displayed activity within the intraparietal sulcus and the dorsal premotor cortex (dPMC). In a recent study that used a similar compensate task, the authors also reported significant activity within these regions (Toyomura et al., 2007). The intraparietal sulcus is involved in spatial transformations for motor preparation (Astafiev et al., 2003; Grefkes et al., 2004; Tanabe et al., 2005). The present findings suggest that the intraparietal
sulcus may be involved not only in somatosensory or visuo-spatial transformations but also in frequency-related transformations as well; this region may be recruited as pitch shifts are transformed into spatial information (i.e., up or down) within the frequency domain. This spatial information may be used by premotor cortices (e.g., dPMC), so that the vocal output can be adjusted in the appropriate direction to correct for the feedback shift. Indeed, recent fMRI studies on motor planning reported that both the intraparietal sulcus and the dPMC may be involved in selecting an appropriate motor response (Beurze et al., 2007; Mars et al., 2007). Moreover, previous studies have associated the dPMC with sensorimotor interactions in general (Chouinard and Paus, 2006; Petrides, 1986), and with auditory–motor interactions in particular (Chen et al., 2006; Zatorre et al., 2007).

In the present study, both groups recruited the right dPMC during the ignore task, while they recruited the left dPMC for the compensate task. Similarly, Toyomura et al.’s (2007) compensate task also resulted in significant activity within the left dPMC. The lateralized recruitment of the dPMC may be explained by the differences in motor planning for each of these tasks. Mars et al.’s (2007) study on motor programming recently reported that the left dPMC is involved in selecting responses due to sensorimotor associations, while the right dPMC is recruited when selecting an alternative to or reprogramming an already-learned response. Our behavioral results suggest that compensating or correcting for pitch-shifted feedback may be more or less automatic for everyone. Therefore, since the compensate task may activate a “default” motor program, activity in the left dPMC may represent the selection of an upward or downward vocal adjustment to counteract the pitch shift. In the ignore task, subjects may have recruited the right dPMC as they tried to keep their vocal output steady in the presence of altered feedback, which required either reprogram-
During both pitch-shifted tasks, both groups recruited the left supramarginal gyrus, which has been implicated in short-term pitch memory (Gaab et al., 2003, 2006). Since subjects needed to remember the target note while either ignoring or correcting for the feedback shift, the increased activity in the supramarginal gyrus may be attributed to maintaining the note in short-term memory. In a similar study, the supramarginal gyrus was also recruited while non-musicians corrected for pitch-shifted auditory feedback; however, increased supramarginal activity was seen only in the right hemisphere, not the left hemisphere (Toyomura et al., 2007).

Comparisons between singers and non-musicians in the ignore task

After statistically comparing the imaging data between both groups, we found that the non-musicians showed greater neural activity within the supramarginal gyrus and the mouth region of M1 during the ignore task. If the supramarginal gyrus is important to maintaining the target note in working memory, non-musicians may need greater access to this system than singers. Non-musicians may have also required greater activity in the mouth region of M1, which may be needed to maintain vocal output in the presence of pitch-shifted feedback. Singers, on the other hand, displayed more activity within auditory cortices (posterior STG and STS) than non-musicians. The posterior STS, which previously has shown a preference for vocal stimuli (Belin et al., 2000; Kriegstein and Giraud, 2004), has also been reported to be involved in processing changes both in vocal source and in spectrotemporal detail, and may have a more specific role in extracting sound features for vocal stimuli, such as pitch height and intensity (Celsis et al., 1999; Warren et al., 2003, 2006). For this task, even though the shifted feedback did not contain the correct note, the singers may have recruited auditory cortices as they used the shifted feedback to monitor their task performance; as long as they kept their vocal output steady, the pitch of the shifted feedback would not vary greatly.

During the ignore task, as well as the compensate task, singers also exhibited greater activity within the putamen than non-musicians. The putamen has been reported to play a role in vocalization—patients with putaminal lesions often exhibit dysarthria or dysphonia (Jurgens, 2002). However, no activity within the putamen was seen during the simple singing task in either group, so the putamen may not have an important role in singing only one note. Putaminal activity, often seen during motor program consolidation, may accompany the recall of a well-learned motor program of correcting inaccurate vocal output (Doyon et al., 2003).

Comparisons between singers and non-musicians in the compensate task

Statistical between-group comparisons of the imaging data determined that non-musicians showed higher activity in the left dPMC, which may have been used to link the pitch-shifted feedback
with a vocal adjustment in the opposite direction of the pitch shift. In contrast, the singers’ extensive vocal experience resulted in increased neural activity within the anterior portion of the rostral cingulate zone (RCZa), the STS, and the putamen. The RCZa is reported to be functionally distinct from the posterior rostral cingulate zone (RCZp) and the caudate cingulate zone (CCZ; Picard and Strick, 2001). The CCZ has been associated with simple motor execution and control, but this area has not been reported to be active during speech (Picard and Strick, 1996). The RCZp has been implicated in voluntary motor selection, which includes speech generation (Paus et al., 1993; Picard and Strick, 1996, 2001). The CCZ has been associated with simple motor execution and control, but this area has not been reported to be active during speech (Picard and Strick, 1996). Therefore, within singers, RCZa activity may have increased as an indicator of high conflict, either between the intended note and the auditory feedback or between two competing motor responses towards the feedback shift. Singers displayed enhanced STS activity in both the ignore and compensate tasks. For both tasks, the singers may have recruited the STS as they analyzed the shifted feedback to monitor their progress in the task.

### Conclusion

Although we found that non-musicians were less accurate at singing single notes than experienced singers, both groups recruited...
similar functional networks for singing, which were comparable to that described in earlier singing studies. Moreover, we have confirmed that the brain regions within this network are functionally connected. In the ignore task, non-musicians could not maintain their vocal output in the presence of pitch-shifted feedback as well as experienced singers; this behavioral difference is accompanied by a difference in neural activity between the groups. Thus, vocal training and experience not only improves voluntary control over the pitch-shift response but may also modulate the functional network for singing during the ignore task. While there were no significant group differences in performing the compensation task, the two groups displayed different neural activity between the groups. Thus, vocal training and experience may also modulate the functional network for singing during the compensate task, each of which may be dependent upon the amount of cognitive conflict, and the STS, suggesting that described in earlier singing studies. Moreover, we have confirmed similar functional networks for singing, which were comparable to that described in earlier singing studies. Moreover, we have confirmed that the brain regions within this network are functionally connected.

Acknowledgments

The authors would like to thank Pierre Ahad, Patrick Bermudez, Marc Bouffard, Vincent Gracco, Talya Grumbach, Henry McDonagh III, and the McConnell Brain Imaging Centre staff for their invaluable technical assistance. This work was supported by grants from the Canadian Institutes of Health Research (CIHR; RJZ), the Eileen Peters McGill Majors Fellowship (JMZ), and the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT; JMZ).

References


Please cite this article as: Zarate, J.M., Zatorre, R.J., Experience-dependent neural substrates involved in vocal pitch regulation during singing. NeuroImage (2008), doi:10.1016/j.neuroimage.2008.01.026


