ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 33 (2006) 628–635

Shared and distinct neural correlates of singing and speaking

Elif Özdemir,[a,b] Andrea Norton,[a] and Gottfried Schlaug[a,*]

[a]*Music and Neuroimaging Laboratory, Beth Israel Deaconess Medical Center/Harvard Medical School, 330 Brookline Ave., Palmer 127, Boston, MA 02215, USA*
[b]*Department of Psychology, Boston University, Boston, MA, USA*

Received 23 February 2006; revised 10 July 2006; accepted 14 July 2006
Available online 7 September 2006

Using a modified sparse temporal sampling fMRI technique, we examined both shared and distinct neural correlates of singing and speaking. In the experimental conditions, 10 right-handed subjects were asked to repeat intoned ("sung") and non-intoned ("spoken") bisyllabic words/phrases that were contrasted with conditions controlling for pitch ("humming") and the basic motor processes associated with vocalization ("vowel production"). Areas of activation common to all tasks included the inferior pre- and post-central gyrus, superior temporal gyrus (STG), and superior temporal sulcus (STS) bilaterally, indicating a large shared network for motor preparation and execution as well as sensory feedback/control for vocal production. The speaking more than vowel-production contrast revealed activation in the inferior frontal gyrus most likely related to motor planning and preparation, in the primary sensorimotor cortex related to motor execution, and the middle and posterior STG/STS related to sensory feedback. The singing more than speaking contrast revealed additional activation in the mid-portions of the STG (more strongly on the right than left) and the most inferior and middle portions of the primary sensorimotor cortex. Our results suggest a bihemispheric network for vocal production regardless of whether the words/phrases were intoned or spoken. Furthermore, singing more than humming ("intoned speaking") showed additional right-lateralized activation of the superior temporal gyrus, inferior central operculum, and inferior frontal gyrus which may offer an explanation for the clinical observation that patients with non-fluent aphasia due to left hemisphere lesions are able to sing the text of a song while they are unable to speak the same words.

Introduction

Recent studies have challenged the classical view of the existence of distinct cerebral modules for music and language processing by showing activation of language specific areas with musical tasks (Binder et al., 1997; Griffiths et al., 1999; Binder et al., 2000; Maess et al., 2001; Koelsch et al., 2002; Gaab et al., 2003; Griffiths, 2003; Patel, 2003; Tillman et al., 2003; Koelsch et al., 2004). In particular, the posterior part of the inferior frontal gyrus commonly thought to contain Broca's area and to be involved in speech production, has been found to be activated by various musical tasks, among them pitch and rhythm discrimination, pitch memory, and musical syntax (Binder et al., 1997; Platel et al., 1997, 1998; Maess et al., 2001; Koelsch et al., 2002, 2005; Griffiths, 2003; Gaab et al., 2003; Patel, 2003). However, this sharing of neural substrates between musical and language tasks conflicts with clinical reports that emphasize a double disassociation between singing and speaking. For instance, it has been reported that patients with Broca's aphasia are able to sing the lyrics of a song better than they can speak the same words (Gerstman, 1964; Geschwind et al., 1968; Geschwind, 1971; Sparks et al., 1974; Yamadori et al., 1977; Brust, 2003). It has been argued that one of the reasons this phenomenon occurs is due to the fact that the left hemisphere is more engaged in propositional speech while the right hemisphere shows greater involvement with automatic or non-propositional speech such as counting or singing familiar songs (Lum and Ellis, 1994).

Brain imaging and lesion studies have shown left hemisphere lateralization for language perception and production (Petersen et al., 1988; Caplan et al., 1995; Caplan et al., 1998; Burton et al., 2000; Wise et al., 2001). On the other hand, some aspects of music such as melody or meter perception, and spectral processing of musical stimuli have been shown to be processed more by the right hemisphere (Zatorre et al., 1994; Binder et al., 1997; Griffiths et al., 1999; Hugdahl et al., 1999; Zatorre and Belin, 2001) while other aspects such as temporal processing of musical stimuli have been shown to be mediated more by the left hemisphere (Peretz, 1990; Belin et al., 1998; Schuppert et al., 2000; Zatorre and Belin, 2001). Overall, music processing may rely more on a bihemispheric network including (but not limited to) the superior temporal gyrus and sulcus, the inferior and superior parietal lobule as well as the inferior frontal gyrus and other parts of the premotor cortex (Schuppert et al., 2000; Gaab et al., 2001; Zatorre and Belin, 2003; Patel, 2005).

* Corresponding author. Fax: +1 617 632 8920.
E-mail address: gschlaug@bidmc.harvard.edu (G. Schlaug).
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2006.07.013

This greater bihemispheric organization for musical function may be the key to understanding why patients with left frontal lesions can sing the lyrics of a song, but cannot speak the words (Gerstman, 1964; Geschwind et al., 1968; Geschwind, 1971; Sparks et al., 1974; Yamadori et al., 1977; Brust, 2003), and thus, could support the hypothesis that there are two possible routes to the articulation of words: (1) a normal language-based route via the left hemisphere, and (2) a singing-based or melodically-intoned route that is either bihemispheric or via the right hemisphere. Some empirical data support a bihemispheric role for the execution and sensorimotor control of vocal production in both singing and speaking (Guenther et al., 1998; Jeffries et al., 2003; Brown et al., 2004) although there seems to be greater left lateralization for speaking under normal physiological conditions. Consistent with these findings are results of studies with Transcranial Magnetic Stimulation (TMS). When applied in an inhibitory mode to the left inferior frontal cortex in normal right-handed subjects, TMS can cause speech interference while the same stimulation to the right homologous region does not interfere significantly with either speech or singing (Epstein et al., 1999; Stewart et al., 2001).

Examining the productive components of music and language presents an obvious challenge for functional imaging studies. Most imaging studies have used covert speech/language tasks in which subjects are instructed to silently recall or express spoken phrases in order to overcome problems inherent in using an overt speaking task in the MR scanner environment (Hinke et al., 1993; Herholz et al., 1996; Ryding et al., 1996; Wildgruber et al., 1996; Ackermann et al., 1998; Friedman et al., 1998; Wildgruber et al., 2001; Perani et al., 2003; Sabbah et al., 2003; Peck et al., 2004). However, several studies have shown that overt and covert speaking tasks actually result in different activation patterns suggesting that covert speech (i.e., suppressed overt speech) might lead to a different or lesser activation pattern, particularly in the sensory feedback regions than that of overt speech (Bookheimer et al., 1995; Barch et al., 1999; Huang et al., 2002; Aziz-Zadeh et al., 2005; Shuster and Lemieux, 2005).

In order to successfully employ an overt vocalization task in the MR scanner environment, several challenges must be overcome. First, jaw and head movements caused by articulation can interfere with image acquisition and may be difficult to correct (Birn et al., 1998). Second, the offline analysis of subjects' recorded responses can be difficult due to the potential interaction between the scanner noise and the responses themselves. For imaging the overt articulation tasks in this study, we modified a sparse temporal sampling method previously used for auditory perceptual experiments (Gaab et al., 2003). In addition to allowing the stimulus presentation and articulated response to occur during the silent period preceding image acquisition, this modified sparse temporal sampling method eliminates the intra- and inter-acquisition effects of scanner noise by having a long repetition time (TR) and a short acquisition time (TA), and by acquiring stacks of axial slices with specific delay times following the subject's vocalization in order to capture the peak hemodynamic response (Hall et al., 1999; Belin et al., 1999; Eden et al., 1999; Talavage et al., 1999; Hall et al., 2000; Gaab et al., 2003).

The major aim of the present study was to examine the shared and distinct neural substrates of overt singing and speaking using the same words/phrases for both conditions. Based on the imaging and lesion studies mentioned above, we hypothesized that there would be a large overlap in the neural representations of singing and speaking with only a few small regions showing distinct activations for either task. Our design improves upon previous studies examining the neural correlates of singing and speaking by (1) using the same bisyllabic words/phrases for both the singing and speaking tasks, (2) using more closely-matched control conditions than previous experiments, and by (3) employing a modified sparse temporal sampling method that allowed us to effectively use production tasks in the scanner and capture subjects' peak hemodynamic response.

Participants and methods

Participants

Ten right-handed volunteers, five males and five females (mean age = 24.2 ± 8.51) without any neurological, psychiatric, or hearing problems participated in this study. None of our subjects were professional singers or were singing in a choir. This study was approved by the Institutional Review Boards of Boston University (BU) and Beth Israel Deaconess Medical Center (BIDMC). All subjects gave written informed consent and received monetary compensation for their participation in the study.

Experimental stimuli and paradigm

A set of 20 bisyllabic words/phrases were recorded by a native speaker of English using Adobe Audition 1.5 software (Adobe, San Jose, CA). The set consisted of words and phrases that are commonly used in everyday life. Recordings were done at a rate of one syllable per second and subjects were instructed to repeat whatever they heard at that same rate. We specifically chose this response rate because we will apply this experimental paradigm to a group of aphasic patients and run between-group comparisons.

In each trial, subjects heard an auditory stimulus followed by a pause of 1.79 s and a short "ding" with a duration of 0.31 s. The "ding" was used as an auditory cue to indicate that the subject should repeat what they had just heard. The pause was inserted prior to the cue in order to increase the delay time between the auditory presentation of the stimuli and the MR acquisition after the subject's response. Considering the timing of this design (see Fig. 1), the actual perception of the words/phrases had little or no effect on the activation elicited by the reproduction of the words/phrases.

Our experimental design consisted of two experimental conditions, intoned (singing) or non-intoned (speaking) bisyllabic words/phrases, and three control conditions. In the speaking condition, subjects heard bisyllabic phrases (e.g., "wake up") or words (e.g., "money") and were asked to repeat exactly what they have heard immediately after the auditory cue. In the singing condition, subjects heard bisyllabic words or phrases melodically intoned on two different pitches (separated by a minor third). In the humming control condition, subjects heard the same two pitches used in the singing condition hummed in either an ascending or descending manner, and were asked to repeat what they had heard immediately after the auditory cue. In the vowel production control condition, subjects heard a spoken vowel pair (e.g., "ah-ah", "oh-oh" as in "go", or "oo-oo" as in "zoo") with minimal intonation or inflection, and were asked to repeat those vowels at the same rate of one vowel per second after the auditory cue. In the silence (rest) condition, subjects heard no stimulus, but were asked to take a breath after the auditory cue to simulate their preparation for the other conditions. Before undergoing fMRI, all subjects spent

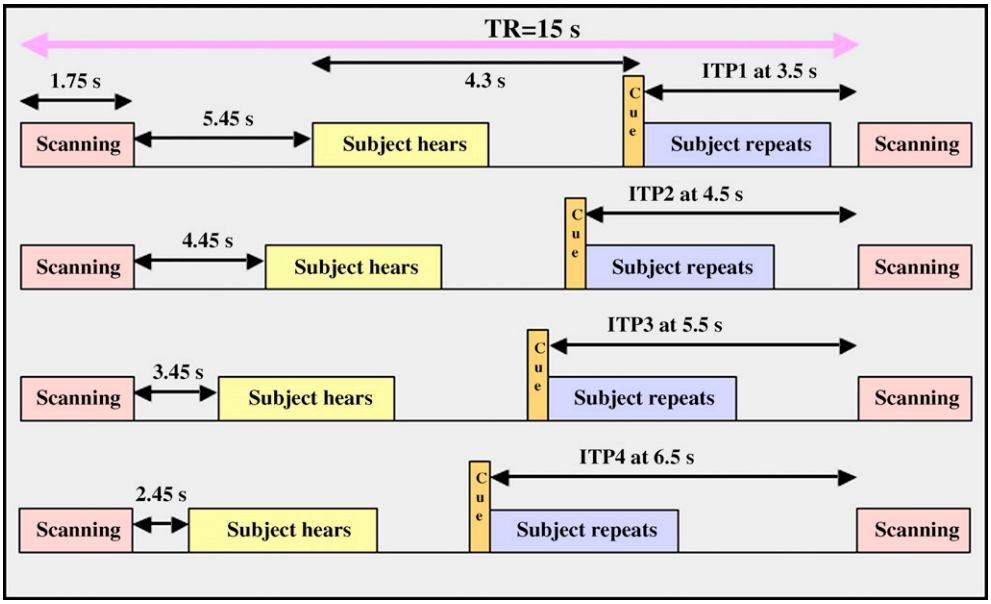

Fig. 1. The delay between subject's responses and the onset of the MR acquisition was varied by shifting the experimental tasks within the 15 s time frame. This jittering resulted in sets of axial images with delays of 3.5, 4.5, 5.5, and 6.5 s for the first image in the stack of axial slices in relation to the auditory cue.

approximately 20 min being familiarized with the stimuli and experimental design and practicing the tasks. Subjects were instructed to keep their eyes closed throughout the imaging session.

fMRI scanning

Functional Magnetic Resonance Imaging (fMRI) was performed on a GE 3.0 T whole-body MRI scanner A gradient-echo EPI-sequence with an effective repetition time (TR) of 15 s, an echo time (TE) of 25 ms, an acquisition time (TA) of 1.75 s, and a matrix of 64×64 was used for functional imaging. Using a midsagittal scout image, a total of 28 axial slices with a voxel size of 3.8×3.8×4 were acquired over 1.75 s after each trial. Initiation of the first set of 28 slices was synchronized with the stimulus presentation using Presentation software version 7.2.6 (Neurobehavioral Systems, Albany, CA). Stimuli were presented binaurally via scanner-compatible headphones. The order in which conditions were presented was randomized across trials. Although the TR was constant at 15 s, the delay between subject's responses and the onset of the MR acquisition was varied by moving the experimental tasks within the 15 s time frame (see Fig. 1). This jittering resulted in sets of axial images with delays of 3.5, 4.5, 5.5, and 6.5 s after the auditory cue for the first image in the stack of axial slices (for more details on this jittering approach, see Gaab et al., 2003). By combining the data from these four jitter points, we were able to capture the peak hemodynamic response within the brain while allowing for timing differences between subjects and brain regions.

Each run was comprised of a total of 20 stimuli (divided equally among the five conditions). Each of the four imaging time points (ITP) was sampled at least 24 times across all runs. Subjects' responses were recorded using a microphone and SoundEdit16 software (Macromedia, San Francisco, CA) for further offline analysis. Offline analyses included verification of subjects' responses after each trial (using the audio recordings from the experiments), an assessment of subjects' delay times in responding after the auditory cue, and an evaluation of subjects' adherence to the required response rate of one syllable per second.

fMRI data analysis

fMRI data were analyzed using the SPM99 software package (Institute of Neurology, London, UK) including realignment, spatial normalization, and smoothing using an isotropic Gaussian kernel (8 mm full-width at half-maximum). Condition and subject effects were estimated according to the general linear model (Friston, 2002). Each scan was scaled in proportion to its global intensity in order to remove the effect of global differences in scan intensity. Low-frequency drifts were removed using a temporal high-pass filter with a cutoff of 179 s (as suggested by the SPM software). A box-car function was applied to the fMRI time series with an epoch length of one. No temporal derivatives were applied. Because of the nature of the sparse temporal sampling design, we did not convolve our data with the hemodynamic response function (HRF) nor apply a low-pass filter (Gaab et al., 2003).

In order to determine whether the peak hemodynamic response was being captured in each subject, a region of interest (ROI) was drawn bilaterally in the inferior precentral gyrus (IPCG) on a spatially standardized anatomical data set using the MRIcro software (version 1.37, Rorden and Brett, 2000). The IPCG was chosen since it had the strongest and most robust activation across all subjects. This ROI (volume of 12.5 cc on either hemisphere) was then superimposed on the beta images of each time point for each subject using the singing vs. silence and speaking vs. silence contrasts. The mean beta values were transformed into z scores separately for each condition and each subject. Fig. 2 shows the hemodynamic response in the inferior precentral gyrus across all subjects for the two experimental conditions. As illustrated by this figure, the delay times allowed us to capture the peak of the hemodynamic response.

By combining all individual imaging time points, a design matrix was modeled to examine the condition effects for each subject

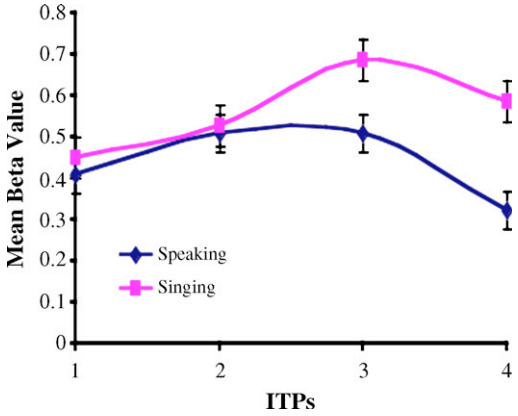

Fig. 2. The ROI analysis of the inferior precentral gyrus showed that the peak hemodynamic response was captured for both singing>silence and speaking>silence conditions with the delayed imaging employed in this experiment.

separately. Each experimental and control condition was contrasted with the silence condition. In addition, the following contrasts were calculated: singing vs. speaking, singing vs. humming, and speaking vs. vowel production for each subject.

The contrast images for each subject were entered into a random effects analysis. One-sample *t* tests were calculated from SPM contrast images for each of the four active conditions (singing, speaking, humming, and vowel production) vs. the silence condition. A threshold of $P<0.05$ (FDR corrected for multiple comparisons) was applied. For the direct comparisons between the active tasks, we applied a mask in order to limit the search volume to the voxels that showed activation in the active vs. silence comparisons. The speaking more than silence contrast was chosen for this mask since it had the largest volume of activation. The mask image was generated from the positive voxels of the speaking more than silence contrast at a threshold of $P<0.005$ (uncorrected). We chose to use this liberal mask in order to include as many voxels as possible in the mask image, then applied it to the paired *t* tests determining significant differences between singing vs. speaking, singing vs. humming, and speaking vs. vowel production comparisons. In order to decrease the possibility of accepting false-positive clusters, only clusters of six or more contiguous voxels were included in the analysis (Forman et al., 1995).

In a preliminary analysis, we compared words with phrases for both the singing and speaking conditions in order to determine whether there would be any differences in the activation pattern. No differences between phrases and words were observed at a threshold $P<0.001$ (uncorrected).

In order to examine the hemispheric laterality effects of the singing vs. speaking contrasts, an ROI was drawn on a spatially standardized anatomical data set using MRIcro representing the superior temporal gyrus on either side (volume of 16.8 cc on each side). This ROI was then superimposed on each subject's SPM *t*-image for the singing vs. speaking contrast. Using a paired *t* test, we examined the mean ROI *t* score to determine whether there was a significant difference between the right and left STG.

Results

Behavioral data

Offline analysis of the subjects' responses recorded during the fMRI experiment indicated that all subjects responded appropriately and used the required response rate. The average response time after the auditory cue was 0.688 (SD=0.189) s.

fMRI data

Contrasting each condition with silence

In a second-level analysis, speaking (Fig. 3a) and singing (Fig. 3b) conditions, each contrasted with the silence condition, showed bilateral activation in the inferior pre- and postcentral gyrus, the inferior frontal gyrus including Brodmann's areas 44 and 45, the middle and posterior portions of the superior temporal

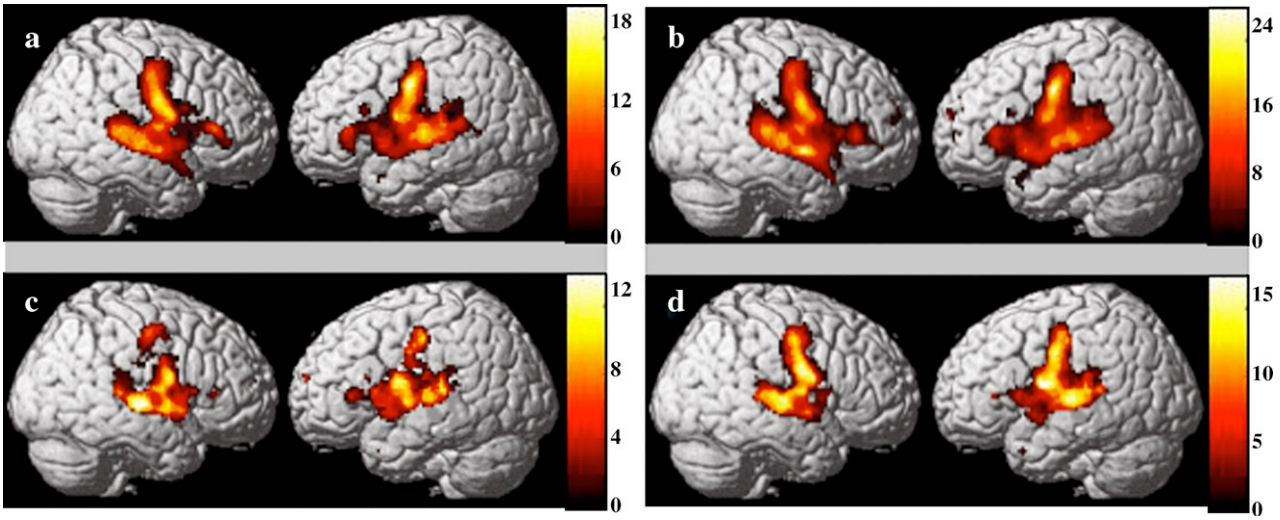

Fig. 3. (a) Speaking>silence contrast ($P<0.05$, FDR corrected, extent threshold=6 voxels). (b) Singing>silence contrast ($P<0.05$, FDR corrected, extent threshold=6 voxels). (c) Humming>silence contrast ($P<0.05$, FDR corrected, extent threshold=6 voxels). (d) Vowel production>silence contrast ($P<0.05$, FDR corrected, extent threshold=6 voxels).

gyrus (STG), and the superior temporal sulcus (STS) ($P<0.05$, FDR corrected). The humming vs. silence contrast (Fig. 3c) showed bilateral activation in the inferior pre- and postcentral gyrus as well as the STG and the IFG ($P<0.05$, FDR corrected). The vowel production vs. silence contrast (Fig. 3d) showed bilateral activation in the inferior pre- and postcentral gyrus and the mid-portion of the STG centered around Heschl's gyrus ($P<0.05$, FDR corrected). The extent of the STG activation was not as pronounced as it was in the speaking and singing conditions.

Comparing singing with speaking

The singing more than speaking contrast (Fig. 4a) showed bilateral activation in the mid-portion of the STG (right significantly stronger than left; $P<0.05$), the most inferior aspect of the central operculum, a small region in the inferior pre- and postcentral gyrus, and the inferior aspect of the inferior frontal gyrus ($P<0.05$, FDR corrected).

Comparing singing with humming

Singing compared to humming (Fig. 4b) showed strong activation of the inferior pre- and postcentral gyrus, the mid-portion of the STG, the STS bilaterally, and anterior portions of the IFG ($P<0.05$, FDR corrected). The singing vs. humming contrast looked similar to the speaking more than silence contrast. Since our singing condition equals speaking with intonation, subtracting the intonation should result in an activation pattern similar to that of the speaking vs silence contrast.

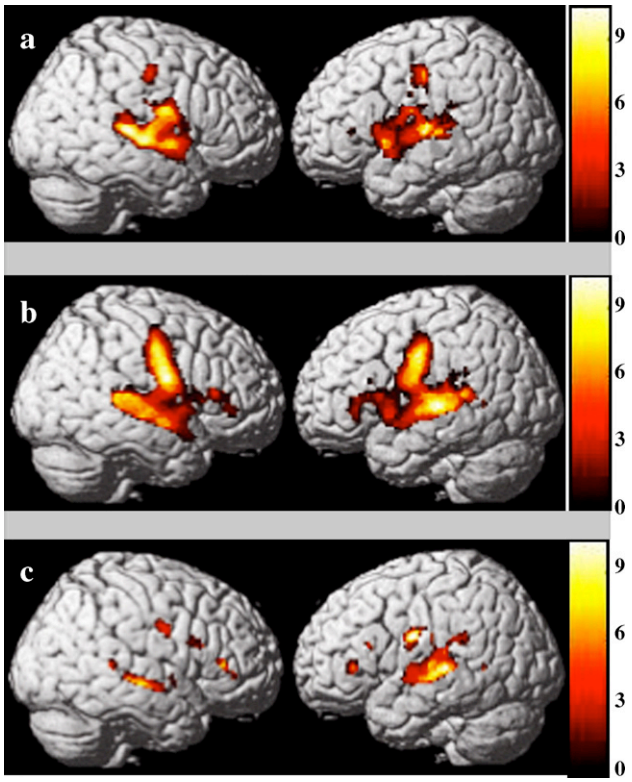

Fig. 4. (a) Singing>speaking contrast ($P<0.05$, FDR corrected, extent threshold=6 voxels). (b) Singing>humming contrast ($P<0.05$, FDR corrected, extent threshold=6 voxels). (c) Speaking>vowel production contrast ($P<0.05$, FDR corrected, extent threshold=6 voxels).

Comparing speaking with vowel production

Speaking compared to vowel production (Fig. 4c) revealed activation in the anterior as well as posterior parts of the inferior frontal gyrus, most likely corresponding to BA 45 and BA 44, the inferior part of the precentral gyrus bilaterally, the left middle and posterior STG, the left STS, and the middle and posterior portions of the right STS ($P<0.05$, FDR corrected).

Discussion

Our results showed a large overlap in activation patterns when subjects sang or spoke bisyllabic words/phrases, articulated vowels or hummed on two pitches. The overlapping regions of activation included the inferior pre- and postcentral gyrus, the superior temporal gyrus, and the superior temporal sulcus on both hemispheres. Some variation across tasks was seen in the inferior frontal gyrus activation which appeared strongest in the singing and speaking tasks, and present, but to a lesser degree, in the humming and vowel production tasks. This pattern of commonly activated brain regions may constitute a shared neural network for the motor preparation, execution, and sensory feedback/control for both intoned and spoken vocal production.

We also found distinct regions of activation when singing was compared with speaking or when speaking and singing were compared with their appropriate control conditions (vowel production and humming, respectively). The singing more than speaking contrast revealed a pattern of activation that was very strong in the anterior to mid-portions of the superior temporal gyrus including Heschl's gyrus, but also in the anterior portion of BA 22 and BA 38. The activation was much stronger on the right than on the left. Shown to be activated in various music studies (Jancke et al., 2002; Gaab et al., 2003; Overy et al., 2004), Brown et al. (2004) found BA 38 to be active in more complex singing tasks compared to monotonic vocalization tasks. Griffiths et al. (1998) found it to be more active with complex musical sounds than with monotonic sequences while Zatorre and Belin (2001) found that the activity in this region covaried with the degree of spectral variation. Furthermore, there is clinical evidence that resection of the anterior temporal lobe in the right hemisphere results in impairments of melody processing (Samson and Zatorre, 1988; Zatorre, 1985; Zatorre and Belin, 2001). Also of interest is the additional activation of the most inferior portion of the inferior frontal gyrus in the singing condition. This could possibly indicate a functional subdivision of Broca's region; the more inferior portion may show greater activation in intoned vocal production conditions than in spoken conditions. The somatotopic relationships of the additional activation of the middle and most inferior portions of the precentral gyrus may indicate that oro-facial and possibly laryngeal movement representations are more activated during singing than during speech.

The speaking more than vowel production contrast revealed a pattern of activation that could represent a rapid articulation network (after controlling for basic motor processes associated with relatively slow vocalization in vowel production tasks) that includes higher order sensorimotor planning and preparatory brain regions in the inferior frontal gyrus, as well as sensory feedback regions in the middle and posterior STG and STS. Accurate perception of one's own vocal output is required for both singing and speaking in order to use feedback information to make necessary adjustments.

The functional activation pattern of the singing more than humming contrast is similar to the pattern of the speaking more than silence contrast. This is not surprising since our experimental design regards singing as "intoned speaking". Thus, subtracting intonation from singing leaves us with the neural correlates for speaking.

Strong bilateral activation found in all four conditions might, at first, seem to contradict the concept of hemispheric lateralization for speaking (more left) and singing (more right) as has been seen in at least one previous study (Riecker et al., 2000). The empirical data that support the notion of hemispheric lateralization for the expressive component of language have been obtained from clinical case studies and neuroimaging studies involving primarily covert language production tasks (Petersen et al., 1988; Caplan et al., 1995, 1998; Burton et al., 2000; Wise et al., 2001). The few studies that have used overt speaking tasks have described a more bihemispheric network involving the pre- and postcentral gyrus as well as the superior temporal gyrus (Brown et al., 2004; Guenther et al., 2006; see also meta-analysis by Indefrey and Levelt, 2004). As evidenced by this meta-analysis study, one of the probable causes for the strong bihemispheric activations seen with our experimental tasks is the overt nature of our tasks and the involvement of both hemispheres in planning, lexical selection, retrieval, self-monitoring, and word production (see Indefrey and Levelt, 2004 for a review on this). A left hemisphere advantage may be more typically seen in situations in which precise and rapid auditory-motor interactions are required (Husain et al., 2005) while right hemisphere advantage might be seen in situations with reduced production speed (articulation rate may be slower in singing than in speaking, potentially reducing the left hemispheric advantage) considering the sustained vowels and syllable chunking that is common in singing. The prosodic features inherent in music (e.g., intonation, change in pitch, syllabic stress) may help speakers chunk syllables into words and words into phrases, and it is possible that this chunking is supported more by right hemisphere structures than by left hemisphere structures.

The activation network (i.e., primary and secondary auditory cortex in the STG, inferior premotor regions, primary sensorimotor regions) that was seen when subjects were asked to repeat intoned (singing condition) and non-intoned (speaking condition) words/phrases is similar to an articulatory network best described by Guenther and his colleagues (Guenther, 2001; Guenther and Ghosh, 2003; Guenther et al., 2006). Based on imaging and cell recording studies, Guenther proposed that three interacting subsystems control speech production: an auditory and a somatosensory feedback subsystem, and a feedforward control subsystem. In this model, the superior temporal gyrus receives projections from the frontal motor cortical areas that predict the sound of one's own voice and compare them with the auditory feedback (Guenther, 2001; Guenther and Ghosh, 2003; Guenther et al., 2006). The somatosensory feedback subsystem consists of primary and higher-order somatosensory areas that encode tactile and proprioceptive information for the sound being produced. As the third component of the model, the feedforward control subsystem involves cortico-cortical projections from premotor to motor cortex (Guenther et al., 2006). Due to the timing of our sparse temporal fMRI method, it is extremely unlikely that the temporal lobe activation was caused by the perception of the auditory stimulus that we asked the subject to repeat. We also tested this by only doing the analysis without the first imaging time point (thus, increasing the time between the end of the auditory stimulus and the first acquisition). There was no difference in the activation of the temporal lobe when all imaging time points were used compared to only using ITPs 2–4 for the analysis. Thus, it is very likely that the temporal lobe activation is related to hearing one's own voice and comparing this with the prediction of the sound which may involve the inferior frontal lobe as well.

Our data support the notion that the majority of sensorimotor processes for singing and speaking are shared (Guenther et al., 1998; Jeffries et al., 2003; Brown et al., 2004). Nevertheless, singing in comparison to speaking activated a larger network that involved regions in the STG (particularly on the right), the inferior aspects of the central operculum and the inferior frontal gyrus. This larger network, in particular, the strong bitemporal lobe and the additional frontal lobe activations, could reflect a greater bihemispheric organization for singing and may also explain why unilateral left frontal lesions affect speaking more severely than singing, assuming that lexical access remains unimpaired (Gerstman, 1964; Geschwind et al., 1968; Geschwind, 1971; Gordon and Bogen, 1974; Sparks et al., 1974; Yamadori et al., 1977; Brust, 2003).

Hickok et al. (2003) investigated neural processes common to both aural perception and covert production for speech and music. The task for speech stimuli consisted of listening to and then covertly rehearsing nonsense sentences while the task for the music stimuli consisted of listening to and then covertly humming melodic tonal sequences. Brain regions found to be active for the listening and production tasks for both speech and music conditions included the left posterior Sylvian fissure at the parietal–temporal boundary (Spt region), left posterior STS, left and right premotor cortex. Additionally, the music condition showed activation in the right posterior STS (Hickok et al., 2003). Thus, these data support an overlap in the perceptual and productive aspects of both speech and music as well as the bihemispheric organization of music production tasks.

From a methodological perspective, our data support the use of a modified sparse temporal sampling method that takes advantage of the inherent delay in the hemodynamic response for imaging production tasks in the scanner. Because of the challenges presented, overt tasks are less commonly used than covert ones (Hinke et al., 1993; Yetkin et al., 1995). In covert language studies, subjects are instructed to silently produce speech; however, there are several disadvantages to this approach. First, a covert task is difficult to control since one cannot verify whether subjects have actually carried out the tasks. Second, the timing of the covert (silent) responses cannot be accurately measured or recorded. Third, since the behavioral, motor, and perceptual components of overt speech differ from those of covert speech, the neural correlates of covert language tasks may also be different from those of overt language tasks (Shuster and Lemieux, 2005). In a TMS experiment, Aziz-Zadeh et al. (2005) showed that overt and covert speech show some hemispheric differences. Stimulation applied over the motor site in the right hemisphere did not effect covert speech, but did impair overt speech.

The tasks used in our study differed from those in previous studies (Perry et al., 1999; Riecker et al., 2000; Jeffries et al., 2003) in that we employed control conditions that were functionally equivalent to our experimental conditions and we controlled the rate at which subjects responded. In order to control for the basic motor processes associated with articulation in the speaking condition, we employed a vowel production task. Similarly, we used a humming condition to control for the low-level sensor-

imotor operations of the singing mechanism and the pitch information present in the singing condition. Thus, singing minus humming should be similar to speaking minus the silence control condition, and the results seen in Figs. 3a and 4c are indeed, very similar. Furthermore, to ensure that the difference in activation patterns observed between the singing and speaking conditions was elicited solely by intonation and not by rhythm or lyrics, we used the same words/phrases and production rate in both conditions, thus allowing for the optimal comparison between the singing and speaking tasks.

Our results suggest a large degree of overlap in neural activation when subjects were asked to overtly repeat sung and spoken words/phrases. The additional activation of the temporal lobe (particularly on the right), central operculum, middle portion of the primary sensorimotor cortex, and inferior portion of the inferior frontal gyrus may hold the key to our understanding of why a clinical disassociation between singing and speaking exists in some patients with Broca's aphasia who can sing the words of a song, but cannot speak them, and why some therapies employing intonation techniques have been reported to facilitate recovery in such non-fluent aphasic patients.

Acknowledgments

We thank Drs. Katie Overy and Nadine Gaab for their help with the design and initial stages of this study. This work was supported in part by NINDS and IFMR.

References

Ackermann, H., Wildgruber, D., Grodd, W., 1998. Does the cerebellum contribute to cognitive aspects of speech production? A functional MRI study in humans. Neurosci. Lett. 247, 187–190.

Aziz-Zadeh, L., Cattaneo, L., Rochat, M., Rizzolatti, G., 2005. Covert speech arrest induced by rTMS over both motor and nonmotor left hemisphere frontal sites. J. Cogn. Neurosci. 17 (6), 928–938.

Barch, D.M., Sabb, F.W., Carter, C.S., Braver, T.S., Noll, D.C., Cohen, J.D., 1999. Overt verbal responding during fMRI scanning: empirical investigations of problems and potential solutions. NeuroImage 10 (6), 642–657.

Belin, P., Zilbovicius, M., Crozier, S., Thivard, L., Fontaine, A., Masure, M.-C., Samson, Y., 1998. Lateralization of speech and auditory temporal processing. J. Cogn. Neurosci. 10, 536–540.

Belin, P., Zatorre, R.J., Hoge, R., Evans, A.C., Pike, B., 1999. Event-related fMRI of the auditory cortex. NeuroImage 10 (4), 417–429.

Binder, J.R., Frost, J.A., Hammeke, T.A., Cox, R.W., Rao, S.M., Prieto, T., 1997. Human brain language areas are identified by functional MRI. J. Neurosci. 17, 353–362.

Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P., Springer, J.A., Kaufman, J.N., Possing, E.T., 2000. Human temporal lobe activation by speech and non-speech sounds. Cereb. Cortex 10, 512–520.

Birn, R.M., Bandettini, P.A., Cox, R.W., Jesmanowicz, A., Shaker, R., 1998. Magnetic field changes in the human brain due to swallowing or speaking. Magn. Reson. Med. 40, 55–60.

Bookheimer, S.Y., Zeffiro, T.A., Blaxton, T., Gaillard, W.D., Theodore, W., 1995. Regional blood flow during object naming and word reading. Hum. Brain Mapp. 3, 93–106.

Brown, S., Martinez, M.J., Hodges, D.A., Fox, P.T., Parsons, L.M., 2004. The song system of the human brain. Brain Res. Cogn. Brain Res. 20 (3), 363–375.

Brust, J.C.M., 2003. The cognitive neuroscience of Music. In: Peretz, I., Zatorre, R. (Eds.), Music and Neurologist: A Historical Perspective. Oxford Univ. Press, Great Britain, pp. 181–191.

Burton, M.W., Small, S.L., Blumstein, S.E., 2000. The role of segmentation in phonological processing: an fMRI investigation. J. Cogn. Neurosci. 12 (4), 679–690.

Caplan, D., Gow, D., Makris, N., 1995. Analysis of lesions by MRI in stroke patients with acoustic–phonetic processing deficits. Neurology 45, 293–298.

Caplan, D., Alpert, N., Waters, G., 1998. Effects of syntactic structure and propositional number on patterns of regional cerebral blood flow. J. Cogn. Neurosci. 10 (4), 541–552.

Eden, G.F., Joseph, J.E., Brown, H.E., Brown, C.P., Zeffiro, T.A., 1999. Utilizing hemodynamic delay and dispersion to detect fMRI signal change without auditory interference: the behavior interleaved gradients technique. Magn. Reson. Med. 41 (1), 13–20.

Epstein, C.M., Meader, K.J., Loring, D.W., Wright, R.J., Weissman, J.D., Shappard, S., Lah, J.J., Puhalovich, F., Gaitan, L., Davey, K.R., 1999. Localization and characterization of speech arrest during transcranial magnetic stimulation. Clin. Neurophys. 110, 1073–1079.

Friston, K.J., 2002. Human brain function, Statistical Parametric Mapping, 2nd ed. Academic Press, Great Britain, pp. 1–74.

Friedman, L., Kenny, J.T., Wise, A.L., Wu, D., Stuve, T.A., Miller, D.A., Jesberger, J.A., Lewin, J.S., 1998. Brain activation during silent word generation evaluated with functional MRI. Brain Lang. 64 (2), 231–256.

Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A., Noll, D.C., 1995. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. Magn. Reson. Med. 33, 636–647.

Gaab, N., Gaser, C., Zaehle, T., Jancke, L., Schlaug, G., 2003. Improvement-related functional plasticity following pitch memory training. NeuroImage 2006 31, 255–263.

Gerstman, H.L., 1964. A case of aphasia. J. Speech Hear. Disord. 29, 89–91.

Geschwind, N., 1971. Current concepts: aphasia. N. Engl. J. Med. 284, 654–656.

Geschwind, N., Quadfasel, F.A., Segarra, J.M., 1968. Isolation of the speech area. Neuropsychologia 6, 327–340.

Gordon, H.W., Bogen, J.E., 1974. Hemispheric lateralization of singing after intracarotid sodium amylobarbitone. J. Neurol. Neurosurg. Psychiatry 37, 727–738.

Griffiths, T.D., 2003. Functional imaging of pitch analysis. Ann. N. Y. Acad. Sci. 999, 40–49.

Griffiths, T.D., Buchel, C., Frackowiak, R.S., Patterson, R.D., 1998. Analysis of temporal structure in sound by the human brain. Nat. Neurosci. 1 (5), 422–427.

Griffiths, T.D., Johnsrude, I.S., Dean, J.L., Green, G.G.R., 1999. A common neural substrate for the analysis of pitch and duration pattern in segmented sound? NeuroReport 10, 3825–3830.

Guenther, F.H., 2001. Neural modeling of speech production. Proceedings of the 4th International Nijmegen Speech Motor Conference.

Guenther, F.H., Ghosh, S.S., 2003. A model of cortical and cerebellar function in speech. Proceedings of the XVth International Congress of Phonetic Sciences, pp. 169–173.

Guenther, F.H., Hampson, M., Johnson, D., 1998. A theoretical investigation of reference frames for the planning of speech movements. Psychol. Rev. 105 (4), 611–633.

Guenther, F.H., Ghosh, S.S., Tourville, J.A., 2006. Neural modeling and imaging of the cortical interactions underlying syllable production. Brain Lang. 96, 280–301.

Hall, D.A., Haggard, M.P., Akeroyd, M.A., Palmer, A.R., Summerfield, A.Q., Elliott, M.P., Gurney, E.M., Bowtell, R.W., 1999. "Sparse" temporal sampling in auditory fMRI. Hum. Brain Mapp. 7, 213–223.

Hall, D.A., Summerfield, A.Q., Goncalves, M.S., Foster, J.R., Palmer, A.R., Bowtell, R.W., 2000. Time-course of the auditory BOLD response to scanner noise. Magn. Reson. Med. 43 (4), 601–606.

Herholz, K., Thiel, A., Wienhard, K., Pietrzyk, U., von Stockhausen, H.-M., Karbe, H., Kessler, J., Bruckbauer, T., Halber, M., Heiss, W.-D, 1996. Individual functional anatomy of verb generation. NeuroImage 3 (1), 185–194.

Hickok, G., Buchsbaum, B., Humphries, C., Muftuler, T., 2003. Auditory-

motor interaction revealed by fMRI: speech, music, and working memory in area Spt. J. Cogn. Neurosci. 15, 673–682.
Hinke, R.M., Hu, X., Stillman, A.E., Kim, S.G., Merkle, H., Salmi, R., Ugurbil, K., 1993. Functional magnetic resonance imaging of Broca's area during internal speech. NeuroReport 4 (6), 675–678.
Huang, J., Carr, T.H., Cao, Y., 2002. Comparing cortical activations for silent and overt speech using event-related fMRI. Hum. Brain Mapp. 15 (1), 39–53.
Hugdahl, K., Bronnick, K., Kyllingsbaek, S., Law, I., Gade, A., Paulson, O.B., 1999. Brain activation during dichotic presentations of consonant-vowel and musical instrument stimuli: a 15O-PET study. Neuropsychologia 37, 431–440.
Husain, F.T., Fromm, S.J., Pursley, R.H., Hosey, L.A., Braun, A.R., Horwitz, B., 2005. Neural bases of categorization of simple speech and nonspeech sounds. Hum. Brain Mapp. (Electronic publication ahead of print).
Indefrey, P., Levelt, W.J., 2004. The spatial and temporal signatures of word production components. Cognition 92 (1–2), 101–144.
Jancke, L., Wustenberg, T., Scheich, H., Heinze, H.J., 2002. Phonetic perception and the temporal cortex. NeuroImage 15 (4), 733–746.
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 (5), 745–749.
Koelsch, S., Gunter, T.C., v Cramon, D.Y., Zysset, S., Lohmann, G., Friederici, A.D., 2002. Bach speaks: a cortical "language-network" serves the processing of music. NeuroImage 17 (2), 956–966.
Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., Friederici, A.D., 2004. Music, language and meaning: brain signatures of semantic processing. Nat. Neurosci. 7 (3), 302–307.
Koelsch, S., Fritz, T., Schulze, K., Alsop, D., Schlaug, G., 2005. Adults and children processing music: an fMRI study. Neuroimage 25, 1068–1076.
Lum, C.C., Ellis, A.W., 1994. Is "nonpropositional" speech preserved in aphasia? Brain Lang. 46 (3), 368–391.
Maess, B., Koelsch, S., Gunter, T.C., Friederici, A.D., 2001. Musical syntax is processed in Broca's area: an MEG study. Nat. Neurosci. 45 (5), 540–545.
Overy, K., Norton, A., Cronin, K., Gaab, N., Alsop, D., Winner, E., Schlaug, G., 2004. Differential lateralization for melody versus rhythm is already present in young children. NeuroReport 15, 1723–1726.
Patel, A.D., 2003. Language, music, syntax and the brain. Nat. Neurosci. 6 (7), 674–681.
Patel, A.D., 2005. The relationship of music to the melody of speech and to syntactic processing disorders in aphasia. Ann. N. Y. Acad. Sci. 1060, 59–70.
Patel, A.D., Gibson, E., Ratner, J., Besson, M., Holcomb, P.J., 1998. Processing syntactic relations in language and music: an event-related potential study. J. Cogn. Neurosci. 10 (6), 717–733.
Peck, K.K., Wierenga, C.E., Moore, A.B., Maher, L.M., Gopinath, K., Gaiefsky, M., Briggs, R.W., Crosson, B., 2004. Comparison of baseline conditions to investigate syntactic production using functional magnetic resonance imaging. NeuroImage 23 (1), 104–110.
Perani, D., Cappa, S.F., Tettamanti, M., Rosa, M., Scifo, P., Miozzo, A., Basso, A., Fazio, F., 2003. A fMRI study of word retrieval in aphasia. Brain Lang. 85 (3), 357–368.
Peretz, I., 1990. Processing of local and global musical information by unilateral brain-damaged patients. Brain 113, 1185–1205.
Perry, D.W., Zatorre, R.J., Petrides, M., Alivisatores, B., Meyer, E., Evans, A.C., 1999. Localization of cerebral activity during simple singing. NeuroReport 10 (16), 3453–3458.
Petersen, S.E., Miezin, F.M., Allman, J.M., 1988. Positron emission tomographic studies of the cortical anatomy of single-word processing. Nature 331, 585–589.
Platel, H., Price, C., Baron, J.C., Wise, R., Lambert, J., Frackowiak, R.S., Lechevalier, B., Eustache, F., 1997. The structural components of music perception. A functional anatomical study. Brain 120, 229–243.
Riecker, A., Ackermann, H., Wildgruber, D., Dogil, G., Grodd, W., 2000. Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. NeuroReport 11 (9), 1997–2000.
Rorden, C., Brett, M., 2000. Stereotaxic display of brain lesions. Behav. Neurol. 12, 191–200.
Ryding, E., BraÅdvik, B., Ingvar, D.H., 1996. Silent speech activates prefrontal cortical regions asymmetrically, as well as speech-related areas in the dominant hemisphere. Brain Lang. 52 (3), 435–451.
Sabbah, P., Chassoux, F., Leveque, C., Landre, E., Baudoin-Chial, S., Devaux, B., Mann, M., Godon-Hardy, S., Nioche, C., Ait-Ameur, A., Sarrazin, J.L., Chodkiewicz, J.P., Cordoliani, Y.S., 2003. Functional MR imaging in assessment of language dominance in epileptic patients. NeuroImage 18 (2), 460–467.
Samson, S., Zatorre, R.J., 1988. Melodic and harmonic discrimination following unilateral cerebral excision. Brain Cogn. 7, 348–360.
Schuppert, M., Munte, T.F., Wieringa, B.M., Altenmuller, E., 2000. Receptive amusia: evidence for cross-hemispheric neural networks underlying music processing strategies. Brain 123 (3), 546–559.
Shuster, L.I., Lemieux, S.K., 2005. An fMRI investigation of covertly and overtly produced mono- and multisyllabic words. Brain Lang. 93 (1), 20–31.
Sparks, R., Helm, N., Albert, M., 1974. Aphasia rehabilitation resulting from melodic intonation therapy. Cortex 10, 303–316.
Stewart, L., Walsh, V., Frith, U., Rothwell, J.C., 2001. Transcranial magnetic stimulation produces speech arrest but not song arrest. Ann. N. Y. Acad. Sci. 930, 433–435.
Talavage, T.M., Edmister, W.B., Ledden, P.J., Weisskoff, R.M., 1999. Quantitative assessment of auditory cortex responses induced by imager acoustic noise. Hum. Brain Mapp. 7 (2), 79–88.
Tillmann, B., Janata, P., Bharucha, J.J., 2003. Activation of the inferior frontal cortex in musical priming. Brain Res. Cogn. Brain Res. 16 (2), 145–161.
Wildgruber, D., Ackermann, H., Klose, U., Kardatzki, B., Grodd, W., 1996. Functional lateralization of speech production at primary motor cortex: a fMRI study. NeuroReport 7, 2791–2795.
Wildgruber, D., Ackermann, H., Grodd, W., 2001. Differential contributions of motor cortex, basal ganglia, and cerebellum to speech motor control: effects of syllable repetition rate evaluated by fMRI. NeuroImage 13 (1), 101–109.
Wise, R.J., Scott, S.K., Blank, S.C., Mummery, C.J., Murphy, K., Warburton, E.A., 2001. Separate neural subsystems within 'Wernicke's area'. Brain 124, 83–95.
Yamadori, A., Osumi, Y., Masuhara, S., Okubo, M., 1977. Preservation of singing in Broca's aphasia. J. Neurol., Neurosurg. Psychiatry 40 (3), 221–224.
Yetkin, F.Z., Hammeke, T.A., Swanson, S.J., Morris, G.L., Mueller, W.M., McAuliffe, T.L., Haughton, V.M., 1995. A comparison of functional MR activation patterns during silent and audible language tasks. AJNR Am. J. Neuroradiol. 16 (5), 1087–1092.
Zatorre, R.J., 1985. Discrimination and recognition of tonal melodies after unilateral cerebral excisions. Neuropsychologia 23 (1), 31–41.
Zatorre, R.J., Belin, P., 2001. Spectral and temporal processing in human auditory cortex. Cereb. Cortex 11 (10), 946–953.
Zatorre, R.J., Evans, A.C., Meyer, E., 1994. Neural mechanisms underlying melodic perception and memory for pitch. J. Neurosci. 14, 1908–1919.