Neural Correlates of Timbre Change in Harmonic Sounds


*Department of Psychiatry and Behavioral Sciences, †Program in Neuroscience, ‡Stanford Brain Research Center, Stanford University School of Medicine, Stanford, California 94305; §Department of Psychology, McGill University, Montréal, Québec, Canada; ‖Institut de Recherche et Coordination Acoustique/Musique (IRCAM-CNRS), Paris, France; and ¶Department of Radiology, Stanford University School of Medicine, Stanford, California 94305

Received October 22, 2001

Timbre is a major structuring force in music and one of the most important and ecologically relevant features of auditory events. We used sound stimuli selected on the basis of previous psychophysiological studies to investigate the neural correlates of timbre perception. Our results indicate that both the left and right hemispheres are involved in timbre processing, challenging the conventional notion that the elementary attributes of musical perception are predominantly lateralized to the right hemisphere. Significant timbre-related brain activation was found in well-defined regions of posterior Heschl’s gyrus and superior temporal sulcus, extending into the circular insular sulcus. Although the extent of activation was not significantly different between left and right hemispheres, temporal lobe activations were significantly posterior in the left, compared to the right, hemisphere, suggesting a functional asymmetry in their respective contributions to timbre processing. The implications of our findings for music processing in particular and auditory processing in general are discussed. © 2002 Elsevier Science (USA)

Key Words: timbre; temporal lobe; superior temporal gyrus; superior temporal sulcus; middle temporal gyrus; fMRI.

INTRODUCTION

An auditory event comprises six general classes of perceptual attributes: pitch, loudness, timbre, perceived duration, spatial location, and reverberant environment (Levitin, 1999). Of these six, timbre is arguably the most important and ecologically relevant feature of auditory events. The timbre of a sound is the principal feature that distinguishes the growl of a lion from the purr of a cat, the crack of thunder from the crash of ocean waves, the voice of a friend from that of a bill collector one is trying to dodge. Timbral discrimination is so acute in humans that most of us can recognize hundreds of different voices and, within a single well-known voice, distinguish whether its possessor is in good health or suffering from a cold—even over the telephone. Much is now known about the psychological representation of timbre (McAdams and Cunibele, 1992; McAdams and Winsberg, 2000; McAdams et al., 1995; Pitt, 1995) but relatively little is known about its neural underpinnings (Shamma, 1999). This report is the first in a series of experiments designed to uncover the functional neuroanatomy of timbre perception in humans and its relation to human psychophysics.

“Timbre” refers to those aspects of sound quality other than pitch, loudness, perceived duration, spatial location, and reverberant environment (McAdams, 1993; Plomp, 1970; Risset and Wessel, 1999). More generally, timbre can be defined as that feature of an auditory stimulus that allows us to distinguish one source from another when all of the other five perceptual features are held constant. That is, timbre is the “tonal color and texture” that allows us to distinguish a trumpet from a clarinet playing the same note or to distinguish one speaker from another when they are reciting the same words. In spite of its prime importance in identification and memory, timbre is one of the least understood components of auditory perception. Helmholtz was the first to recognize that it is to some extent related to a tone’s spectral composition (Helmholtz, 1863/1954). More recent investigators have found that it is crucially related not just to spectral parameters but also to the temporal aspects of tone, such as the rate of increase in level or the presence of inharmonic components in the initial part of the tone (Grey, 1977), acoustic features that affect the “attack quality” (Hajda et al., 1997). For example, a piano is easily recognized as such in a taped recording, but if that same recording is played backwards, the distinct timbre that identifies “piano” is lost, even though the long-term frequency spectrum remains unchanged for any given tone (Berger, 1964).

Accurate memory for the timbral qualities of everyday music was shown recently in a compelling study by Schellenberg and his colleagues (1999). Participants in this study were able to identify popular songs from brief excerpts of only 100 and 200 ms. The authors argued that this was too brief an excerpt for cues such as rhythm, pitch, or tonality to trigger memory and that it must have been timbral memory that was being accessed. Timbral memory may also be the mechanism by which some absolute pitch possessors identify tones (Takeuchi and Hulse, 1993). The distinctive sound of a
The psychological space for timbre representation has been investigated in an extensive series of studies by McAdams et al. (1995) and his colleagues, among others. Since the work of Plomp (1970), the principal paradigm for uncovering the perceptual structure of timbre has employed multidimensional scaling of dissimilarity judgments among sets of musical sounds equated for pitch, loudness, and duration and generally presented over headphones or loudspeakers (see McAdams, 1993; Hajda et al., 1997, for reviews). Over many different studies, two or three perceptual dimensions are generally extracted. The acoustic correlates of these “timbre spaces” depend on the set of sounds tested and to some extent the listener population. For example, using a set of synthesized sounds designed to imitate orchestral instruments or hybrids of these instruments, McAdams et al. (1995) found a three-dimensional space with specific features on several of the timbres (Fig. 1). The acoustic correlates of the commonly shared dimensions were the logarithm of the attack time (the time it takes the energy envelope to go from perceptual threshold to its maximum), the spectral center of gravity, or spectral centroid (a measure of the relative presence of high-frequency versus low-frequency energy in the frequency spectrum), and spectral flux (a measure of how much the spectral envelope changes over the duration of a tone).

Although the psychoacoustics of timbre continues to be extensively investigated (McAdams et al., 1995), information about its neural substrates remains sparse. In general, lesion studies have shown that the right temporal lobe is involved in spectral processing, i.e., pitch matching, frequency discrimination (Milner, 1962; Robin et al. 1990; Sidtis and Volpe, 1988; Zatorre, 1988), and the left temporal lobe is involved in temporal processing (Efron, 1963; Robin et al., 1990). Samson and Zatorre (1994) examined timbre discrimination in patients with unilateral temporal lobe excisions, right and left, while manipulating both the spectral and temporal components of timbre. To their surprise, only those subjects with right temporal lobe excisions were impaired in timbre discrimination, independently of the temporal and spectral cues. These findings support a functional role for the right temporal lobe in timbre discrimination, but leave ambiguous the role of the left temporal lobe, if any.

A few neuroimaging studies have examined the neural basis of music perception in general, with attempts to factionate musical processing into component operations. The emerging picture is that music perception per se does not have a specific neural locus, but that the components of music perception—specifically pitch and rhythm—may. One early PET study showed a specialized role for the right temporal lobe in pitch perception (Zatorre et al., 1993); a more recent PET study by Griffiths et al. (1999) showed remarkably similar bilateral activation for both pitch and duration, in the cerebellum, posterior superior temporal cortices, and inferior frontal cortices. A PET study by Blood et al. (1999) examined affective processing of musical pieces and found activation of neocortical and paralimbic structures in response to pleasant/unpleasant musical stimuli. Belin et al. (2000) found regions in the superior temporal sulcus (bilaterally) that were involved in the processing of human voices. Although the study did not explicitly address timbre perception, timbre-related attributes would have been among the primary cues for subjects in this experiment to distinguish human from environmental sounds and to distinguish human voices from one another.

The functional neuroanatomy of timbre perception in humans remains almost entirely unknown, and yet this important topic may hold the key to understanding fundamental questions about voice identification, musical processing, and memory for the numerous sound sources encountered in everyday life. The only known study to date identified right superior and middle frontal gyri as candidates for selective attending to timbre in the presence of simultaneous variation in pitch and rhythm (Platel et al., 1997), yet this finding is difficult to interpret for several reasons: first, the timbral change was subtle and primarily spectral in nature (a “bright” vs “dull” oboe) and second, the data analysis employed a subtraction in which selective attending to timbral change was compared to selective attending to pitch and rhythm changes. Johsru et al. (1997) examined processing of acoustic transients in nonspeech sounds using PET.

FIG. 1. The timbre space for 18 synthetic musical instrument sounds found by McAdams et al. (1995). A three-dimensional space with acoustic correlates of attack time, spectral centroid, and spectral flux was found. For comparison, Timbres A and B from the present study are shown approximately as they would appear in the space on the basis of these same acoustic parameters. Note that the change from Timbre A to Timbre B occurs as a diagonal across the space since all three dimensions change.

piano or saxophone—presumably based on its timbre—is believed to serve as a cue for pitch identification among absolute pitch possessors who are unable to identify the pitch of sine tones or of instruments with which they are less familiar.

The functional neuroanatomy of timbre perception in humans remains almost entirely unknown, and yet this important topic may hold the key to understanding fundamental questions about voice identification, musical processing, and memory for the numerous sound sources encountered in everyday life. The only known study to date identified right superior and middle frontal gyri as candidates for selective attending to timbre in the presence of simultaneous variation in pitch and rhythm (Platel et al., 1997), yet this finding is difficult to interpret for several reasons: first, the timbral change was subtle and primarily spectral in nature (a “bright” vs “dull” oboe) and second, the data analysis employed a subtraction in which selective attending to timbral change was compared to selective attending to pitch and rhythm changes. Johsru et al. (1997) examined processing of acoustic transients in nonspeech sounds using PET.
This study is of particular interest because psychophysical and cognitive studies of timbre have shown transients to be an important component in timbral identification and recognition (Berger, 1964). Participants in this study discriminated pure tone stimuli incorporating frequency glides of either long (100 ms) or short (30 ms) duration. Comparison of these two conditions revealed significant activation foci in the left orbitofrontal cortex, left fusiform gyrus, and right cerebellum. The left hemisphere activations were consistent with previously cited work implicating left hemisphere structures for temporal processing (Efron, 1963; Robin et al., 1990).

Based therefore on psychophysical models of timbre perception, which point to an important role for both spectral and temporal parameters of sound, we hypothesized a critical role for the left and right temporal lobes in tasks involving timbral processing, as well as those left hemisphere regions previously implicated by Johnsrude et al.'s (1997) study of transient perception. In order to maximize the difference along acoustic dimensions known to affect timbre perception, we selected two types of sounds that differed maximally in terms of attack time, spectral centroid, and spectral flux, as found by McAdams et al. (1995). The two distinct timbre stimuli (A and B) created for the present experiment differed significantly along these three dimensions. Their approximate positions with respect to the original space are shown in Fig. 1. Our experimental design was as follows. We compared brain activation in response to melodies composed of tones with a fast attack, low spectral centroid, and no spectral flux (Timbre A) to that resulting from melodies formed of tones with a slow attack, higher spectral centroid, and greater spectral flux (Timbre B). The theoretical interest in these sounds is that they include spectral, temporal, and spectrotemporal parameters that have been shown to be correlated with perceptual dimensions of timbre. A case in which timbres changed across the entire perceptual space needed to be tested first in order to demonstrate that perceptually salient stimuli reveal differences in activation as measured with fMRI. The critical question is thus: what are the characteristics of brain responses associated with differences in melodic stimuli created with precisely controlled synthetic sounds having different timbres?

**MATERIALS AND METHODS**

**Subjects**

Ten healthy adult subjects (4 males and 6 females; ages 21–45, mean 33) participated in the study after giving written informed consent, and all protocols were approved by the human subjects committee at Stanford University School of Medicine. All participants were volunteers and were treated in accordance with the APA “Ethical Principles of Psychologists and Code of Conduct.”

**Stimuli**

All sounds were created digitally using a 22,050-Hz sampling rate and 16-bit resolution, although stimulus presentation in the MRI scanner used only the 8 most significant bits. Each sound had an amplitude envelope defined in terms of linear attack time (AT), sustain time (ST), and exponential decay time (DT). These three values were, respectively, 15, 60, and 198 ms for Timbre A and 80, 0, and 117 ms for Timbre B (Fig. 2). The main interest was in attack time and the other two values were adjusted to achieve similar perceived durations across the two timbres. The sounds were composed of harmonic spectra. The levels of the harmonics decreased as a power function of harmonic rank ($A_n = n^4$). The exponent of this function was used to establish the base spectral centroid (SCbase), which is defined as the amplitude-weighted mean harmonic rank. For a pure tone SCbase = 1, whereas for a tone composed of the first three harmonics with amplitudes equal to the inverse of the harmonic rank, SCbase = $(1^4 + 2^4 + 3^4)/(1 + 2 + 3) = 1.63$. Expressed in this way, the relative amplitudes among harmonics remain constant with variations in fundamental frequency. Spectral flux ($\Delta$SC, expressed as a change in amplitude-weighted mean harmonic rank) was defined as a variation over time of SC with a rise time SCRT and a fall time SCFT. Between time 0 and SCRT, SC rose from SCbase to SCbase + $\Delta$SC, following a raised cosine trajectory. From time SCRT to SCFT + SCFT, SC decreased back to SCbase, following a slower raised cosine trajectory. After time SCFT + SCFT, SC remained constant at SCbase. Timbre A was characterized by a fast attack (AT = 15 ms), a low base spectral centroid (SCbase = 1.05), and no spectral flux ($\Delta$SC = 0). Timbre B was characterized by a slower attack (AT = 80 ms), a higher base spectral centroid (SCbase = 1.59), and a higher degree of spectral flux ($\Delta$SC = 4.32, SCRT = 49 ms, SCFT = 102). For Timbre B, the instantaneous SC therefore varied from 1.59 to 5.91 between 0 and 49 ms and from 5.91 back to 1.59 between 49 and 151 ms, remaining at 1.59 thereafter (Fig. 2). Prior to the experimental session, the two timbre stimuli were matched psychophysically for loudness by four raters, and this necessitated attenuating the sounds with Timbre B by 8.5 dB compared to those with Timbre A.

Stimuli consisted of 60 six-note melodies played with these timbres. The range of fundamental frequencies was 288.2 Hz (C No. 4) to 493.9 Hz (B4). One of these melodies, realized with each of the timbres, is shown in a spectrographic representation in Fig. 2. Ten melodies with a given timbre constituted a block. All melodies were tonal and ended on the tonic note. The pitch range was confined to an octave but varied from five to nine semitones across melodies; 21 melodies had a range with five semitones (ST), 21 with seven ST, and 18 with nine ST. Each block had 3 or 4 five-ST melodies, 3 or 4 seven-ST melodies, and 3 nine-ST melodies. Thirty melodies had one contour change, and 30 had two changes (five of each per block). There were 36 major-mode and 24 minor-mode melodies (6 major/4 minor per group). Key changes from melody to melody in order to keep the mean pitch constant across melodies. An attempt was made to keep the pitch distribution relatively constant across blocks. The time interval between onsets of successive tones within a melody was 300 ms, making the total duration of each melody approximately 1.8 s for Timbre A melodies and 1.7 s for Timbre B melodies. A new melody occurred every 2.4 s and a block of 10 melodies thus lasted 24 s.
fMRI Tasks

Each subject performed two runs of the same experiment in the scanner. Each experiment consisted of 12 alternating 24-s epochs (epochs of 10 melodies with a given timbre), for a total of 120 melodies per scan. There were 6 Timbre A epochs interleaved with 6 Timbre B epochs. Each of the six blocks of ten melodies was presented for each timbre, but the order of blocks and the order of melodies within blocks were randomized for each subject. Subjects were asked to engage their attention on the melodies. For each six-note melody they were asked to press a response key as soon as the melody was over. In one run, the first epoch was played with Timbre A, and in the second run, the first epoch was played with Timbre B.

Procedure

The order of the two runs was randomized across subjects. They were briefed on the task they were to perform during the scan and practiced the experiment prior to entering the scanner.

Behavioral Data Analysis

The aim of the behavioral data collection was to verify that listeners were paying attention to the stimuli to the same extent for A and B epochs. Reaction time (RT) and number of correct and incorrect responses for A and B epochs were collapsed across the two runs. A response was considered correct if the subject pressed the appropriate response key after the end of the musical segment, but before the onset of the next stimulus. The percentages of correct responses and RTs to correct trials were compared using two-tailed paired t tests. The significance threshold was set at 0.05.

fMRI Acquisition

Images were acquired on a 3T GE Signa scanner using a standard GE whole head coil (software Lx 8.3). A custom-
built head holder was used to prevent head movement; 25 axial slices (5 mm thick, 0 mm skip) parallel to the AC/PC line and covering the whole brain were imaged with a temporal resolution of 1.6 s using a T2*-weighted gradient echo spiral pulse sequence (TR, 1600 ms; TE, 30 ms; flip angle, 70°; 180 time frames; and one interleave) (Glover and Lai, 1998). The field of view was 220 × 220 mm², and the matrix size was 64 × 64, providing an in-plane spatial resolution of 3.44 mm. To reduce blurring and signal loss arising from field inhomogeneities, an automated high-order shimming method based on spiral acquisitions was used before acquiring functional MRI scans (Kim et al., 2000). Images were reconstructed, by gridding interpolation and inverse Fourier transform, for each of the 180 time points into 64 × 64 × 25 image matrices (voxel size, 3.44 × 3.44 × 5 mm). A linear shim correction was applied separately for each slice during reconstruction using a magnetic field map acquired automatically by the pulse sequence at the beginning of the scan (Glover and Lai, 1998).

To aid in localization of functional data, a high-resolution T1-weighted spoiled gradient recalled (SPGR) inversion recovery 3D MRI sequence was acquired with the following parameters: TI, 300 ms; TR, 8 ms; TE, 3.6 ms; flip angle, 15°; 22-cm field of view; 124 slices in the sagittal plane; 256 × 192 matrix; two averages; acquired resolution, 1.5 × 0.9 × 1.1 mm. The images were reconstructed as a 124 × 256 × 256 matrix with a 1.5 × 0.9 × 0.9 mm spatial resolution. Structural and functional images were acquired in the same scan session.

Stimulus Presentation

The task was programmed using Psyscope (Cohen et al., 1993) on a Macintosh (Cupertino, CA) computer. Initiation of scan and task was synchronized using a TTL pulse delivered to the scanner timing microprocessor board from a “CMU Button Box” microprocessor (http://poppy.psy.cmu.edu/psyscope) connected to the Macintosh. Auditory stimuli were presented binaurally using a custom-built magnet compatible system. This pneumatic delivery system was constructed from a flexible plastic tubing attached to a cone-shaped funnel, which in turn was connected to flexible plastic tubing leading to the participant’s ears. This delivery system filters the sound with an overall low-pass characteristic and a spectral dip in the region of 1.5 kHz (Fig. 3). The tubing passed through foam earplug inserts that attenuated external sound by approximately 28 dB. The sound pressure level at the head of the participant due to the fMRI equipment during scanning was approximately 98 dB (A), and so after the attenuation provided by the ear inserts, background noise was approximately 70 dB (A) at the ears of the listener. The experimenters set the stimuli at a comfortable listening level determined individually by each participant during a test scan.

Image Preprocessing

fMRI data were preprocessed using SPM99 (http://www.fil.ion.ucl.ac.uk/spm). Images were corrected for movement using least squares minimization without higher order corrections for spin history (Friston et al., 1996) and were normalized to stereotaxic Talairach coordinates using nonlinear transformations (Ashburner and Friston, 1999; Friston et al., 1996). Images were then resampled every 2 mm using sinc interpolation and smoothed with a 4-mm Gaussian kernel to reduce spatial noise.

Statistical Analysis

Statistical analysis was performed on individual and group data using the general linear model and the theory of Gaussian random fields as implemented in SPM99. This method takes advantage of multivariate regression analysis and corrects for temporal and spatial autocorrelations in the fMRI data (Friston et al., 1995). Activation foci were superimposed on high-resolution T1-weighted images and their locations were interpreted using known neuroanatomical landmarks (Duvenoy et al., 1999). MNI coordinates were transformed to Talairach coordinates using a nonlinear transformation (Brett, 2000).

A within-subjects procedure was used to model all the effects of interest for each subject. Individual subject models were identical across subjects (i.e., a balanced design was used). Confounding effects of fluctuations in the global mean were removed by proportional scaling where, for each time point, each voxel was scaled by the global mean at that time point. Low-frequency noise was removed with a high-pass filter (0.5 cycles/min) applied to the fMRI time series at each voxel. A temporal smoothing function (Gaussian kernel corresponding to a half-width of 4 s) was applied to the fMRI time series to enhance the temporal signal to noise ratio. Regressors were modeled with a boxcar function corresponding to the epochs during which each timbre was presented and convolved with a hemodynamic response function. We then defined the effects of interest for each subject with the relevant contrasts of the parameter estimates. Group analysis was performed using a random-effects model that incorporated a two-stage hierarchical procedure. This model estimates the error variance for each condition of interest across subjects, rather than across scans, and therefore provides a stronger generalization to the population from which data are acquired (Holmes and Friston, 1998). In the first stage, contrast images for each subject and each effect of interest were generated as described above. In the second stage, these contrast images were analyzed using a general linear model to determine voxelwise t statistics. One contrast image was generated per subject, for each effect of interest. A one-way, two-tailed t test was then used to determine group activation for each effect. Finally, the t statistics were normalized to Z scores, and significant clusters of activation were determined using the joint expected probability distribution of height and extent of Z scores (Poline et al., 1997), with height (Z > 1.67; P < 0.05) and extent (P < 0.05) thresholds.

Contrast images were calculated using a within-subjects design for the following comparisons: (1) Timbre B–Timbre A; (2) Timbre A–Timbre B.
Behavioral

End-of-melody detections did not differ in accuracy between A and B blocks (t(8) = -0.900; P = 0.40), nor did RT (t(8) = -1.876; P = 0.10). Accuracy rates were 99.5 ± 2.8% for Timbre A and 99.8 ± 1.0% for Timbre B. RTs were 1914 ± 59.5 and 1931 ± 76.0 ms for Timbres A and B, respectively. We conclude that there were no behavioral differences (and thus no confounding attentional or arousal differences) between the two sets of stimulus blocks containing the two timbres.
Brain Regions That Showed Significant Activation Differences between the Two Timbre Stimuli

<table>
<thead>
<tr>
<th>Timbre B–Timbre A</th>
<th>P value (corrected)</th>
<th>No. of voxels</th>
<th>Z score</th>
<th>Peak Talairach coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right STG/STS (BA 41/22/21)</td>
<td>&lt;0.001</td>
<td>297</td>
<td>3.31</td>
<td>50, –8, –4</td>
</tr>
<tr>
<td>Left STG/STS (BA 41/22/21)</td>
<td>&lt;0.002</td>
<td>285</td>
<td>2.55</td>
<td>–36, –24, 8</td>
</tr>
<tr>
<td>Timbre A–Timbre B</td>
<td>No significant clusters</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Note. For each significant cluster, region of activation, significance level, number of activated voxels, maximum Z score, and location of peak in Talairach coordinates are shown. Each cluster was significant after height and extent correction (P < 0.05). STG, superior temporal gyrus; MTG, middle temporal gyrus; BA, Brodmann Area.

**TABLE 1**

Brain Activation

Timbre B–Timbre A. Two significant, tightly focused, clusters of activation were observed in both the left and right temporal lobes (Table 1, Fig. 4). In both hemispheres, the main activation foci were localized to the dorsomedial surface of the superior temporal gyrus (STG) (BA 41/42/22), adjoining the insular cortex and extending to the upper and lower banks of the STS (BA 21). Additionally, in the left hemisphere, a more posterior and medial activation focus overlying the circular insular sulcus was also observed.

Seven of the nine subjects showed significant activation in the left temporal lobe cluster and eight also showed activation in the right temporal lobe cluster. There was no difference overall between the percentages of voxels activated in left and right temporal lobes (Wilcoxon matched pairs test, Z = 0.296; P = 0.76). Similarly, no difference was seen in the average intensity of activation between the two hemispheres (Z = 1.006; P = 0.31). Interestingly, activation in the left temporal lobe was posterior to the right temporal lobe activation (Fig. 5). In the left hemisphere, the most anterior activation point was at –34, –20, 8 and the most posterior point was at –41, –40, 10. In the right hemisphere, the most anterior point was at 54, –4, –8, and the most posterior point was at 40, –28, –6.

To quantify this asymmetry in activation, we examined the number of voxels activated anterior and posterior to y = –22 (midway between the anteriormost and posteriormost activations). Only 14 voxels were activated in the left anterior region compared to 271 in the left posterior region, whereas 286 voxels were activated in the right anterior region compared to only 11 in the right posterior region. Three of the nine subjects had no activated voxels in the left anterior region and six had no activation in the right posterior region.

To further analyze the asymmetry in activation, we examined the Hemisphere by Anterior/Posterior interaction using Analysis of Variance (ANOVA). For this analysis, we constructed identical and (left/right) mirror symmetric regions of interest (ROI) in the left and right hemispheres based on the observed group activation. Within these ROIs, the number of voxels activated in each Hemisphere in the Anterior and Posterior subdivisions was computed for each subject. As shown in Fig. 6, ANOVA revealed a significant Hemisphere x Anterior/Posterior interaction (F(1,8) = 20.11; P < 0.002). Mean hemispheric differences were not significant (F(1,8) = 0.21; P = 0.66), nor were mean Anterior/Posterior differences (F(1,8) = 0.72; P < 0.42).

Timbre A–Timbre B. No brain regions showed greater activation during Timbre A compared to Timbre B.

**DISCUSSION**

Our findings strongly support the notion that timbre-related information is processed in both the left and right hemispheres. Furthermore, timbre-related differences in activation were detected only in mid sections of the STG, STS, and adjoining insular cortex. Although the overall extent of activation was not significantly different in the left and right temporal lobes, left temporal lobe activations were significantly posterior to right temporal lobe activations, suggesting a functional asymmetry in the left and right temporal lobe regions involved in timbre processing. The present study provides information about the precise temporal lobe regions involved in timbre processing. The implications of these findings for timbre processing are discussed below.

Specifically, we found that timbre stimuli with greater spectral flux, higher spectral centroid, and slower attack (Timbre B) resulted in greater activation than timbre stimuli with low spectral centroid, no spectral flux, and faster attack (Timbre A). No brain regions showed significantly greater activation during Timbre A compared to Timbre B. Another functional imaging study of timbre processing identified the right superior and middle frontal gyri as brain regions involved in timbral processing (Platel et al., 1997), a finding that was based on a comparison of selective attending to timbral change with selective attending to pitch or rhythm change rather than a direct comparison of two distinct timbre stimuli. The present study did not find significant activation in these regions. The lack of prefrontal cortical activation in the present study may be related to the fact that subjects did not have to make any complex comparisons or decisions about the timbre stimuli, nor did the task require the involvement of selective attentional processes. Unlike the Platel et al. study, which used a subtle, primarily spectral timbral change (“bright” vs “dull” oboe), the present study used timbre stimuli that have been shown to be perceptually very dissimilar (McAdams et al., 1995). The subtle timbral changes may also have been the reason for the lack of temporal lobe activation in the Platel et al. study. Our data are in
agreement with those of lesion studies which have suggested involvement of the temporal lobe in timbre processing (Milner, 1962; Samson and Zatorre, 1994). Lesion studies from the Montreal Neurological Institute have reported that right temporal lobe lesions result in greater impairment in timbre discrimination than left temporal lobe lesions (Milner, 1962; Samson and Zatorre, 1994). Nevertheless, a closer look at the published data also suggests that left temporal lesions do impair timbre discrimination. These studies have not, however, provided any details of the precise temporal lobe regions involved in timbre perception and have instead focused on examining the relative contributions of the entire left and right temporal lobes to timbre processing. The present study clearly indicates that focal regions of the left and right temporal lobes are involved in timbre processing.

Equal levels of activation were detected in the left and right temporal lobes. To our knowledge, lateralization of timbre-related activation has never been directly tested before. Within the temporal lobes, activation was localized to the primary auditory cortex (PAC) as well as the auditory association areas in the belt areas of the STG and in the STS. Anteriorly, the activation foci did not extend beyond Heschl’s gyrus (HG) (Howard et al., 2000; Kaas and Hackett, 2000; Tian et al., 2001). The most anterior activation point was at y = −4 mm (Talairach coordinates) in the right temporal lobe. The bilateral activation observed in the present study is in agreement with human electrophysiological studies that have found bilateral N100 sources during timbre-specific processing (Crummer et al., 1994; J ones et al., 1998; Pantev et al., 2001).

Interestingly, activation for Timbre B was greater than that for Timbre A, despite the matching of the two stimuli for loudness. One plausible explanation is that Timbre B has a greater spectral spread (or perceivable spectral bandwidth) than Timbre A and would thus have a more extensive tonotopic representation, at least in the primary and secondary auditory cortices (Howard et al., 2000; Wessinger et al., 2001). While the two timbres have the same number of harmonics and thus ideally would have the same bandwidths, it is likely that the higher harmonics of Timbre A are below the auditory threshold and/or masked by the scanner noise, thus creating a perceptible bandwidth difference. Bandwidth differences were also present in the Wessinger et al. (2000) and Hall et al. (2001) studies. However, in all three cases, other factors were also varied (the relative periodicity of the signals in Wessinger et al., the pitch in Hall et al., and attack time and spectral flux in the present study), so the true contribution of bandwidth per se cannot be unambiguously determined. Another possibility is that the greater spectral flux in Timbre B may result in greater cortical activation, particularly in the STS, as this region appears to be sensitive to temporal changes in spectral distribution. Finally, the relative salience of the stimuli may also be important since Timbre B is clearly more salient perceptually than A, but it is not easy to measure “salience” per se. At any rate, it may not be possible to independently control salience and timbre, since some timbral attributes, such as brightness and sharp attack, generally seem to make sounds “stand out.”

Our results also bear upon the cortical representation of loudness and its examination using fMRI since the fMRI activation patterns of two loudness-matched stimuli are so unbalanced. Studies of the effect of sound level on the extent and level of temporal lobe activation have shown complex relationships based on the nature of pure/harmonic-complex tones used (Hall et al., 2001) as well as the precise subregion and laterality of the auditory cortical activation examined (Brechmann et al., 2002). For example, sound level had a small, but significant, effect on the extent of auditory cortical activation for the pure tone, but not for the harmonic-complex tone, while it had a significant effect on the response magnitude for both types of stimuli (Hall et al., 2001). Furthermore, the relationships among sound level, loudness perception, and intensity coding in the auditory system are complex (Moore, 1995). In any case, our results discount simple models for loudness that are based, implicitly or explicitly, on the degree of activation of neural populations.

Although there were no overall hemispheric differences in extent of activation, our results clearly point to differences in the specific temporal lobe regions activated in the left and right hemispheres. ANOVA revealed that distinct subregions of the STG/STS were activated in each hemisphere: activation in the left hemisphere was significantly posterior to right hemisphere activation, with very few voxels showing activation in the anterior regions (y > −22) in the left hemisphere. Conversely, very few voxels showed activation in the posterior regions (y < −22) in the right hemisphere. These results point to a hemispheric asymmetry in neural response to timbre; indeed, the location of peak activation in left auditory cortex was 16 mm posterior to that of peak activation in right auditory cortex. This shift is 8–11 mm more posterior than can be accounted for by morphological differences in the locations of the left and right auditory areas (Rademacher et al., 2001). Based on a detailed cytoarchitectonic study of postmortem brains, Rademacher et al. have reported a statistically significant asymmetry in the location of the PAC in the left and right hemispheres. The right PAC is shifted anteriorly by 7 or 8 mm and laterally by 4 or 5 mm, with respect to the left PAC. An almost identical pattern of functional asymmetry was observed in the present study—the right hemisphere activation peak was shifted more anteriorly and laterally than the activation peak in the left hemisphere. While the centroid of the right PAC is roughly shifted by ΔX = 4, ΔY = 8, and ΔZ = 1 mm, the peak fMRI activations in the present study are shifted by ΔX = 14, ΔY = 16, and ΔZ = −12 mm in the X, Y, and Z directions, respectively. Thus, even after accounting for the anatomical asymmetry in location of the PAC, left hemisphere activation is still considerably posterior to right hemisphere activation, thus suggesting a functional asymmetry in the left and right temporal lobe contributions to timbre processing. Furthermore, these hemispheric differences do not appear to arise from volumetric differences since careful measurements of the highly convoluted auditory cortex surface have revealed no size difference between the left and right PAC or HG (Rademacher et al., 2001). Moreover, the functional asymmetry observed during timbre processing extends beyond the dorsal STG to the STS and the adjoining circular insular sulcus as well.
In each hemisphere, two distinct subclusters of activation could be clearly identified, one in the STG and the other localized to the STS. The anterior-most aspects of the dorsomedial STG activations in both hemispheres were localized to a stripe-like pattern of activation in area T1b which follows the rostromedial slope of HG (Brechmann et al., 2002). In the left hemisphere, the activation extends posteriorly to area T2, which is centered on Heschl’s sulcus. In both hemispheres, the activations appear to be anterior to area T3, which principally overlaps the main body of the planum temporale. As defined by Brechmann et al. (2002), area T1b borders the circular sulcus medially, the first transverse sulcus anteriorly, and the roof of HG caudolaterally. The caudally adjacent T2 was centered on Heschl’s sulcus, thus including the posterior rim of HG and a similar area behind the sulcus on the anterior planum temporale/second Heschl’s gyrus. There is very little activation in area T1a, the division of the HG rostral to the first transverse sulcus. Areas T1a and T1b overlap to a great extent the PAC probability maps and Talairach coordinate ranges reported by Rademacher et al. (2001). Based on the landmarks identified by Rademacher et al. and Brechmann et al., it therefore appears that in the
left hemisphere, the dorsal STG activation extends posteriorly beyond the PAC to the posterior HG/anterior tip of the planum temporale (BA 42/22) in the belt areas of the auditory association cortex. In the right STG, activation was primarily observed in anterior and posterior HG, including the PAC (BA 41). No activation was observed in the auditory association cortex A2 covering the planum polare (BA 52) anterior to the primary auditory cortex in either the left or right STG.

Although processing of nonlinguistic, nonvocal, auditory stimuli is generally thought to take place in the STG, we also found significant activation in the STS and adjoining circular insular sulcus and insula in both hemispheres. The STS regions activated in the present study have direct projections to the STG (Pandya, 1995) and are considered part of the hierarchical organization of the auditory association cortex (Kaas and Hackett, 2000). These STS zones receive input primarily from primary and secondary auditory cortices in the STG, and the upper bank receives input exclusively from the STG (Seltzer and Pandya, 1978). Note that these STS regions are anterior to, and functionally distinct from, the visual association areas of the STS that are involved in visual motion and gaze perception. This central STS region is not generally known to be activated by pure tones or even by frequency modulation (FM) of tones, in contrast to the T1b and T2 regions of the auditory cortex (Brechmann et al., 2002). Relatively little is known about the auditory functions of the STS, but Romanski et al. (1999) have proposed the existence of a ventral temporal lobe stream that plays a role in identifying coherent auditory objects, analogous to the ventral visual pathway involved in object recognition. Interestingly, Belin et al. (2000) have shown that similar central STS auditory areas are activated when subjects listen pas-
sively to vocal sounds, whether speech or nonspeech. The present study suggests that these STS regions can be activated even when stimuli do not have any vocal characteristics. A common feature that the vocal sounds used in the Belin et al. study and the Timbre B stimuli used in the present study have is that both involve rapid temporal changes in spectral energy distribution (spectral flux). We suggest that timbre, which plays an important role in both music and language, may invoke common neural processing mechanisms in the STS. Unlike the STG, the precise functional divisions of the left and right STS are poorly understood. One clue that may, in the future, aid in better understanding the functional architecture of STS is that we observed a functional hemispheric asymmetry in the STS similar to that in the STG, in that left STS activations were also posterior to those observed in the right STS. The posterior insular regions activated in this study are known to be directly connected to the STG (Galaburda and Pandya, 1983), suggesting a role in auditory function, but the role of this region in auditory processing is less well understood than that of the STS.

The differential contributions of the left and right STG/STS to timbre processing are not clear. Lesion studies have provided very little information on this issue, partly because of the large extent of the excisions (Ligeeois-Chauvel et al., 1998). It is possible that the right hemisphere bias for timbre processing observed in lesion studies may arise partly because more posterior left STG regions that contribute to timbre processing are generally spared in temporal lobectomies. One important reason that the contribution of the left temporal lobe to timbre processing is of considerable interest is the fact that the same acoustical cues involved in the perception of musical timbre may also be involved in processing linguistic cues. For example, spectral shape features may provide a more complete set of acoustic correlates for vowel identity than do formants (Zahorian and Jagharghi, 1993), analogous to the manner in which timbre is most closely associated with the distribution of acoustic energy across frequency (i.e., the shape of the power spectrum). Evidence for this comes from Quesnel's study in which subjects were easily taught to identify subtle changes in musical timbre based on similarities between these timbres and certain English vowels (Quesnel, 2001). Although it has been commonly assumed that the right and left hemispheres play dominant roles in musical and language processing, respectively (Kolb and Wishaw, 2000), our findings suggest that music and language processing may in fact share common neural substrates. In this regard, our data are consistent with human electrophysiological recordings which have also found equivalent bilateral STG activation during music and language processing (Creutzfeldt and Ojemann, 1989). It has also been suggested that the left STG may be critically involved in making distinctions between voiced and unvoiced consonants (Ligeeois-Chauvel et al., 1999). Conversely, the right STG is involved in processing vocal sounds, whether speech or non-speech, compared to nonvocal environmental sounds (Belin et al., 2000).

The present study marks an important step in charting out the neural bases of timbre perception, a critical structuring force in music (McAdams, 1999), as well as in the identifica-

FIG. 6. To characterize the observed hemispheric asymmetry in temporal lobe activation, we used analysis of variance (ANOVA) on the dependent variable “percentage of voxels activated,” with factors Hemisphere (Left/Right) and Anterior/Posterior divisions. ANOVA revealed a significant Hemisphere × Anterior/Posterior interaction (F(1,8) = 20.11; P < 0.002), with the left hemisphere activations significantly posterior to the right hemisphere activations. The coronal slice at y = −22 mm (Talairach coordinates), midway between the anterior- and posteriormost activation in the temporal lobes, was used to demarcate the anterior and posterior subdivisions. Note that this analysis was based on activation maps obtained in individual subjects.
tion of sound sources in general (McAdams, 1993). This study motivates thinking about acoustic stimuli in terms of their timbral properties. This, in turn, will help to establish a more ecological framework for interpreting brain activations arising from acoustic stimuli with different timbral properties. Second, it provides a review of neuropsychological investigations of timbre perception that is integrated with a detailed description of the hemispheric asymmetries in the differential response to sounds that fall within different regions of timbral space.

Future studies will need to examine the relative contribution of the different components of timbre with independent, parametric control. Further electrophysiological studies are also needed that examine how the spectral profile shape and its dynamics are encoded by neurons in the auditory cortex and how the tonotopic and temporal patterns of neuronal activity relate to timbre perception (Shamma, 1999). We expect that such future work will allow us to examine the neural basis of some of the known perceptual structure of timbral space.

This research was supported by NIH Grant HD 40761 and a grant from the Norris Foundation to V.M., by NSERC Grant 228175-00, FCAR Grant 2001-SC-70936, and CFI New Opportunities Grant 2719 to D.J.L., and NIH Grant RR 09784 to G.H.G. We are grateful to Nicole Dolensky for her assistance with the behavioral data analysis.

ACKNOWLEDGMENTS

REFERENCES


